

10

Db 251453 CTATAATAAGACCCAGGTAGCTGTACAGGAGATCTGTATACACATCATTCCTGATCAACA 251394

tracheo

The present sequence represents the complete genome of *Chlamydia trachomatis*. Open reading frames (ORFs) of the genome encode

Align seg 1/1 to: RAD20238 from: 1 to: 1799

```

1 MetArgLysSerValGlyLeuThrIleLeuSerLeuSe 17
101 ATGGCAAGATATACAGTGGAAATCTGTATACCACTCTCTTAGCCTCTC 150
17 vAluValLeuGlnGlyCysLysGluSerHisSerSerThrSerArg 34
151 CGTAGTCTCTCAAGGCTGCAAGAGTCCAGTCACTCTCTATCATCTCGG 200
34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
201 GAGAACTCGCTATTATATATATATATATATATATATATATATATATAT 250
51 GlnValArgLeuLeuSerGluLeuSerLeuValLysHisIleIleGlu 67
251 CAAGTGGGACTTCTTTCAGAAATCAGCCCTCTCAAAATATATCTATGAGG 300
67 yLeuValGlnGlnAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
301 ATTAGTTCAGAAATATATATATATATATATATATATATATATATATAT 350
84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
351 AAGACTACTCTCTCTCTCGGAGCGACTCACTTATATCTTTTAAAGTAAA 400
101 SerAlaPheIlePheSerAspGlyProLeuThrAlaGluAspPheIle 117
401 TCAGCTTTTGGAGTAAATGGGACCCCTTAAACGCTGAAGACTTTATAGA 450
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
451 ATCTTGGAAACAGTAGTACTCTCAGAAAGTCTCAGAAATCTAGCTTTG 500
134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
501 CCTTGAATCCAAATTAATAATGTACGAAAGATCCAAAGAGGACACCTCTCC 550
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValTh 167
551 ATAGACCAATTTGGAGTGCACCTCTCTAATGAATCTACACTTGTGTAC 600
167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184
601 CTGGAATTCCTCCACCTCGGACTTTCTTAAACCTTTTACCTCTCCAGTCT 650
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
651 TTTTCCCGGTTCATAAATCTCAAAAGACCTGCAATCCAAATCTCTACCT 700
201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrp 217
701 ATAGCAAGCGGAGCTTTCTATCTAAATAATATCAAAACAAACAAATGGAT 750
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThr 234
751 AAAACTCTCAAAACAAACCTCCTACTATATCAAGATCAGGTGGAAACATA 800
234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
801 AAACGATTACGATTCACTTCAATTCCTCGATGCAACACAGCAGCAAAACTA 850
251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgIle 267
851 TTTAATCAGGGAACCAATCAATTCGCAAGACCTCTCTGGGAGAGACGCAT 900
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
901 TCTCTAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACTTACACTCTT 950
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300

```

```

951 TTGATGTCGCGAGAACCTCATGTGCTCACCTTCAATATCAATAAATCCCC 1000
301 LeuAsnAsnMetLysLeuArgGluAlaIleAlaSerAlaLeuAspLysG 317
1001 CTCACCAATATGAAGCTTAGAGAACCTTAGCATCAGCTTATAGTAAGGA 1050
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHis 334
1051 AGCTCTTGTCTCACTATATCTTAGGCCGTGCAAAACTGCCGATCATC 1100
334 euLeuProThrAsnIleHisSerTyrProGlnHisGlnLysGlnGluMet 350
1101 TCCTACCTACAAATATTCATAGCTATCCGCAACATCAAAACCAAGAGATG 1150
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGlu 367
1151 GCACAGCCCAAGCTTACGCTTAAAAACTCTTTAAAGAACTTTAGAAGA 1200
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePhePro 384
1201 ACTCAAAATCACTGCTTAAGATCTCGAATCTTATCTTATCTTCCCG 1250
384 alSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluInTrp 400
1251 TTTCTCTGTCAGCAAGTCTTTACTAGTCCAACTTATACGAGACAGTGG 1300
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAla 417
1301 AAGAAGATTTAGGTTCTGATCTCTATCTCTATGTCGGAAGAAATTTGCTCT 1350
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
1351 TCTCAAGCAGACCTATCTCTCAGGGAACCTCTCTTAGCTACAGAGGAT 1400
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAla 450
1401 GGTTCGCGAGACTTTGCTGATCTCTATGCTGCACTTCTAAGCATCTTGTCTTAT 1450
451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGlu 467
1451 CATCAGAGTCTCTCTTATGCAATCAACCAATAGGACTTCTTAGAAAT 1500
467 eLeuGlnAsnIleGluGlnGlnAspHisGlnLysArgSerGluLeu 484
1501 TCTCAAAACATAGACAGACAGACAGATCACCACAAACGCTCGGAATTAG 1550
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleLeuGluPro 500
1551 TGTGCAAGCTTCTCTTACCTAGAGACTTTCATATATTAGGCGGATC 1600
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeu 517
1601 TACCACAGCAGCTTCAATTTCTATGAATAAAACTTTCTAATCTAGG 1650
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
1651 AGTCTCACCACAGAGGTTGTGACTTCCGTATGCTAGGAGAAAT 1696

```

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA91990

seq_documentation_block:

ID AA91990 standard; DNA; 1230025 BP.

XX AA91990;

XX 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; ss.

XX

Chlamydia pneumoniae.

W09927105-A2.

03-JUN-1999.

20-NOV-1998; 98WO-IB01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(G8ST) GENSET.

Griffais R;

WPI; 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Claim 1; Page 291-611; 1912pp; English.

The present sequence represents the complete genome of Chlamydia pneumoniae, and encodes proteins AY34584-135879. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AY34584-135879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.

Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

alignment_scores:

Quality: 2726.00 Length: 533
Ratio: 5.124 Gaps: 1
Percent Similarity: 99.812 Percent Identity: 99.812

alignment_block:

US-09-824-567-2 x AAX91990 ..

Align seg 1/1 to: AAX91990 from: 1 to: 1230025

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17

244480 ATCGCAAGATATCATAGTGGGAATCTGTATCACCATTCTCCCTAGCCTCTC 244529

17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34

244530 CGTAGCTCTCCAGGCTCGAAGGAGTCCAGTCACTCTCTATCATCTCGGG 244579

34 lylGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50

244580 GAGAACTCGCTATTATATAGAGATGACCCCGCTCTTTAGTCCCAAGA 244629

51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGlu 67

244630 CAAGTCGCACTCTCTTCAGAAATCAGCTGTGTCAACATATCTATGAGGG 244679

67 YLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAla 84

244680 ATTAGTTCAGAAATAATATCTTTCAGGAATATAGAGCTCTCTTCAG 244729

84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100

244730 AAGACTACTCTCTTCTCCGAGGACTCCTATACCTTTTAACTGAAA 244779

101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG 117

244780 TCAGCTTTTGGAGTAATGCGACCCCTTAACAGCTGAGACNCTTTATAGA 244829

117 uSerTrpLysGlnValAlaThrClnGluValSerGlyIleTyrAlaPheA 134
244830 ATCTTGGAAACAGTAGTACTCAAGAAGTCTCAGGAATCTATGCTTTTG 244879

134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
244880 CCTTGATCCCAATTAAATATGTACAAAGATCCAGAGGACACCTCTCC 244929

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
244930 ATAGACCATTTTGGAGTGCACCTCTCTTAATGAATCTACACTTGTGTTAC 244979

167 rLeuGlnSerProThrSerHisPheLysLeuLeuAlaLeuProValP 184
244980 CCTGGAATCCCCAACCTCGCATTTCTTAAACTTTTAGCTCTTCCAGTCT 245029

184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
245030 TTTTCCCGCTTCATAANTCTCAAGAACCCTGCAATCCAAATCTTACCT 245079

201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpIl 217
245080 ATAGCAAGCGAGCTTCTATCTCTAAATAATCAACAAACAAATGGAT 245129

217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
245130 AAAACTCTCAAAAACCCCTCACTACTATATAACAAAGTCAGGTGGAACCTA 245179

234 ystThrIleThrIleHisPheIleProAspAlaAsnThrAlaLysLeu 250
245180 AAGGATTAGGATTCACTTCAATCCCGATCGAACAACACAGCAGCAAACTA 245229

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyValArgI 267
245230 TTTAATCAGGAAACTCAATGTCAGAGGACCTCTTGGGAGAACGAT 245279

267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
245280 TCCTCAGAAACCCCTATCCAAATTTACAGTCTAAGGGGCACTTACACTCTT 245329

284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
245330 TTGATGTCCAGGAACTCATGGCTCACTTCAATATCAATAAATTCCTC 245379

301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGl 317
245380 CTCACAAATATGAGCTTAGAAGCCTTAGCATCAGCCTTAGATAGGA 245429

317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
245430 AGCTCTTGTCTCACTATATTTCTAGGCGCTGCATAAACTGCCGATCATC 245479

334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
245480 TCCTACTACAAATATTCATAGCTATCCCGACACACCAAAACAGAGATG 245529

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluGl 367
245530 GCACAAACCCCAAGCTTACGCTAAAAAATCTTTAAAGAGCTTTAGAGA 245579

367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
245580 ACTCCAAATACCCCTAAGATCTGAACATCTTAACTTATCTTATCTTCCGG 245629

384 alSerSerAlaSerSerLeuValGlnLeuIleArgGluGlnTrp 400
245630 TTTTCTCTCAGCAAGTCTTTACTAGTCCAACTTTATCGAGAACAGTGG 245679

401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
245680 AAGAAAGATTAGGTTTCGCTATCTCTTATTCGGAAGGAATTCCTCT 245729

417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434

```

|||||
245730 TCTCAGCAGACCTATCTTCAGGAACTCTCTTTAGCTACAGGAGAT 245779
434 tpPheAlaaspPheAlaaspProMetaIaPheLeuThrIlePheAlaTyr 450
|||||
245780 GGTTCGCAGACTTGTGTATCTCTATGATGATTTCTTAAGCATCTTGTCTAT 245829
451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI 467
|||||
245830 CCATCAGAGTCTCTCTATGCAATCAACCATTAAGACTTCTTAGAAT 245879
467 eLeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuV 484
|||||
245880 TCTACAAACATAGACAGAGCAAGATCACAAAAACGCTCGGAATTAG 245929
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
|||||
245930 TGTCCAAAGCTCTCTTACCTAGAGACTTTCATATATTATGACGATC 245979
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLysLeuSer AsnLeuG 517
|||||
245980 TACCACGACGATTCATTTGCTATGATTAATAAAACTTTCATATCTAG 246029
517 lyValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
|||||
246030 GAGTCTCACCACAGGAGTGTGGACTTCGGTTATGCTAAGGAAAT 246076

```

seq_name: /SIDS5/gcdata/geneseq/geneseq-emb1/NA1999.DAT-AAZ01425

seq_documentation_block:

ID AAZ01425 standard; DNA; 1038602 BP.

AC AAZ01425;

DT 07-OCT-1999 (first entry)

DE Complete genome sequence of Chlamydia trachomatis.

XX Vaccine; eye disease; conjunctivitis; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

XX Chlamydia trachomatis.

XX W09928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX Griffiths R;

XX WFI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Claim 1; Page 373-656; 1755pp; English.

XX The present sequence represents the complete genome of Chlamydia
 CC trachomatis. Open reading frames (ORFs) of the genome encode
 CC polypeptides AA36754-f37949. The polypeptides can be used as vaccines
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 CC be used to control growth of the microorganism. Chlamydia trachomatis is
 CC responsible for a large number of diseases, e.g. eye diseases such as
 CC conjunctival trachoma, nonendemic trachoma, paratrachoma, and inclusion
 CC conjunctivitis; genital diseases such as nongonococcal urethritis,

CC epididymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
 CC pneumopathy in breast feeding infants; and venereal
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in
 CC treating these diseases.

XX
 SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

alignment_scores:

Quality: 1356.00 Length: 539

Ratio: 3.146 Gaps: 10

Percent Similarity: 79.963 Percent Identity: 50.278

alignment_block:

US-09-824-567-2 x AAZ01425/rev ..

Align seg 1/1 to reverse of: AAZ01425 from: 1 to: 1038602

1 MetArgylsleSerValGlyIleCysIleThrIleLeuLeuSerIeuSe 17
 252097 ATGCGCAAGATATACGTGGAACTCTGCTGCTAGCATTAGCAACTC 252048
 17 rValValLeuGlnGlyCysLysGluSerSerHisSerSerThrSerArg 34
 252047 T.....GGATGTTCAAAATCTCTCTAAGCAACCCATCGGT 252010
 34 ly.....GluLeuAlaIleAsnIleArgAspGluProArgSer 46
 252009 CTCACGACTACTCACACAGTGTGCTGAAGCTAAGATGATCTCGACA 251960
 47 LeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuValIysH 63
 251959 TTTGATCTCGAGAGGTTCGCTTCTTTCTGATATCAATTGATCATCA 251910
 63 sIleTyrGluGlyLeuValGlnGluAsnLeuSerGlyAsnIleGluP 80
 251909 TCTCTATGAGGATTTGTATCAAGAACTCCT...TCTGGAGAGTCTTCC 251863
 80 rAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThr 96
 251862 CTGCTTACGCGAGGTTCCTTATCCGAGATATAAAACTTATCT 251813
 97 PheLysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAlaG 113
 251812 TTCACCTTGAANAAGCTTTTGGAGCAATGGAGATCTTATTACCGCTCA 251763
 113 uAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSerGly 130
 251762 TGATTTTCTGCTTCTGGAATGATGTGTACAAAATCGTCTCGCTAGTA 251713
 130 leTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGlnGlu 146
 251712 TTTATCTTTTCGCTTCTCCCTATT...GACGTGAATAAG..... 251675
 147 GlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSerTh 163
 251674GATTCTGGATTTTTCACAAAGATGATCATAC 251643
 163 rLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeuLeu 180
 251642 TCTTGTATCAATCTCTCACTCCAACTCCACATTTTCTAAAGCTGCTTA 251593
 180 laLeuProValPhePheProValHisLysSerGlnArgThrLeu..... 194
 251592 CCTCTCCGCTATTATTTATCTCTGTCAT...TCGACATCATCATATGAAA 251546
 195 GlnSerLysSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnI 211
 251545 GAAGAAAATCTCTCCGATPATCTACTGGAGCTTTTCTTAAAGAGAA 251496
 211 eLysGlnLysGlnThrIleLysSerLysAsnProHisTyrTyrAsnG 228
 251495 GAAGACCGAGATGGTTAAGCTAGAGAGAGCCCTTACTACTATAATA 251446

```

alignment_scores:
  Quality: 148.50      Length: 89
  Ratio: 2.216        Gaps: 3
  Percent Similarity: 75.281  Percent Identity: 38.202

alignment_block:
  US-09-824-567-2 x AQ990639 ..

Align seg 1/1 to: AQ990639 from: 1 to: 538

36  LeuAlaIleAsnNileAArgAspGluProArgSerLeuAspProArgGlnVa 52
    |||: ||| :|||:|||||:|||||:|||||:|||||:|||||:
246  TTAGTCCGTAATAACGCTCTGACCGCAATCGTTAGATCCATCAAGAT 295

52  IargLeuLeuSerGluIleSerLeuValIysHisIleTyGluGlyLeuV 69
    |||: :||| :|||:|||||:|||||:|||||:|||||:|||||:
296  TGAAGGATGCGCAATCATCATTTGGCCCGTGAATCTGTTCAAGGCTTG 345

69  aIGInGluAsnAsnLeuSerGlyIAsnIleGluProAlaLeuAlaGluAsp 85
    |||: :|||:|||||:|||||:|||||:|||||:|||||:
346  TCATCTGTTGGCCG...AATGCGAAATTTTGGCAGGTTCTGCGACGAT 392

```

Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssr@bath.ac.uk
This is one of 2,112 random reads from the M13 library. For

```

seq_name: gb_gss:BH370951

seq_documentation_block:
LOCUS      BH370951
DEFINITION AG-ND-119H23.TF ND-TAM Anopheles gambiae genomic clone AG-ND-119H23.TF
            , DNA sequence.
ACCESSION  BH370951
VERSION    BH370951.1 GI:17317076
KEYWORDS   GSS
SOURCE     African malaria mosquito.
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE  1 (bases 1 to 593)
            Shetty J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
            Direct Submission of BAC-end sequences from Anopheles gambiae
            Unpublished (2001)
OTHER_GSS: AG-ND-119H23.TF
COMMENT

```

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igrr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center, University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For
Seq primer: M13 Rev
Class: BAC ends.

```

FEATURES
  source
    1..593
      /organism="Anopheles gambiae"
      /strain="PEST"
      /db_xref="taxon:7165"
      /clone="AG-ND-119H23"
      /clone_lib="ND-TAM"
      /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      97 a 180 c 171 g 145 t
ORIGIN

alignment_scores
  Quality: 137.00      Length: 159
  Ratio: 1.384        Gaps: 6
  Percent Similarity: 62.264      Percent Identity: 29.560

alignment_block
  US-09-824-567-2 x BH370951/rev

Align seq 1/1 to reverse of: BH370951 from: 1 to: 593

291 TrpLeuThrPheAsnIle..AsnLysPheProLeuAsnAsnMetLysLeuA 307
576 TATCTGGCGTTCAACACCAAGAGAAAGCGTTTCGACACCGTGTGGTAC 527
307 rgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSerThrIle 323
526 GCCAGGCTCTGAGCATCGCTGGACAAAGAGCCATCGTTCCGCGGTA 477
324 PheLeuGlyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHi 340
476 TTCAAGACACAGCGCACCGCGCTATTTCGATCTCCCGCGGCGATGCT 427
340 sSerTyrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrA 357
426 GGGCTATAACGAC...AAACTGCCCGAATACGCTTACACCCGCGAAAG 380
357 lalLysLysLeuPheLysGluAla...LeuGluGluLeuGlnIleThrAla 372
379 CGGCTGAACCTGCTCAAGCAGCGCGCGCTGGAGAAAGGGTTTGAGACC... 333
373 LysAspLeuGluHisLeuAsnLeuIlePheProValSerSerAlaSe 389
332 ...GATATCTGTCGATCGCGTAGCCCGCCCTTACACCCGAAT...TC 289
389 rSerLeuLeuValGlnLeuLeuArgGluGlnIntrLysGluSerLeuGlyP 406
288 ACGCCCATCATCGTCAGATGCTCAAAACGACTGGCGGAA...GTGGCG 242
406 heAlaIleProIleValGlyGluPheAlaLeuLeuGlnAlaAspLeu 422
241 TAAAGCCAAATCGTCACCTCGGAATGGGCCAGTATCTGGCGGGTGG 192
423 SerSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAl 439

```

```

:::||||: ... ||| ||||||:||||:
191 CGCAAGGGGAACAGCAGCGCGCTGTATGCTGTGATCGGATACGG 142
439 aAspProMetAlaPheLeuThrIle 447
::||||| ||| |||:
141 CGACCGGACAACTTCGTACCCCTG 117

seq_name: gb_gss:B07758

seq_documentation_block:
LOCUS      B07758          601 bp      DNA      linear      GSS 15-NOV-1997
DEFINITION      8219P101E01120195T7 Rhodobacter sphaeroides 2.4.1 genomic DNA
                  library to Rhodobacter sphaeroides genomic clone 8219P101E01120195T7
                  similar to dppA (P23847), DNA sequence.
ACCESSION      B07758
VERSION        B07758.1 GI:2055550
KEYWORDS       GSS.
SOURCE         Rhodobacter sphaeroides.
ORGANISM       Rhodobacter sphaeroides
                Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
                Rhodobacter.
REFERENCE      1 (bases 1 to 601)
AUTHORS       Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., Weinstock,G.M.
                and Kaplan,S.
TITLE          Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
                Chromosome II is a true chromosome
JOURNAL        Microbiology 143, 3085-3099 (1997)
MEDLINE        98015398
COMMENT        Contact: Mackenzie, C.
                Department of Microbiology and Molecular Genetics
                The University of Texas Health Science Center
                6431 Fannin St., Houston, TX 77006, USA
                Tel: (713) 794-1742
                Fax: (713) 794-1782
                Email: mackenzie@uth.tmc.edu
                Seq primer: pBluescript SK (-) T7
                Class: shotgun.

FEATURES
  source
    1..601
      /organism="Rhodobacter sphaeroides"
      /strain="2.4.1T"
      /db_xref="taxon:1063"
      /clone="8219P101E01120195T7"
      /clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA
      library"
      /lab_host="E. coli SL7-1"
      /note="Vector: pLA2917; Genomic DNA from Rhodobacter
      sphaeroides was prepared and partially digested with
      Sau3AI. Size selected (20kb) fragments were subcloned into
      the BglII site of the cosmid vector pLA2917 (Allen, L. N.
      and R. S. Hanson, 1985. Journal of Bacteriology,
      161:955-962. The library was then ordered around
      chromosome II of Rhodobacter sphaeroides (Choudhary et
      al., 1994. Journal of Bacteriology, 176:7694-7702). The
      cosmids were then digested with a variety of restriction
      enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and
      DNaseI) and restriction fragments subcloned into the
      respective multiple cloning site sites of pBluescript SK
      (-). Note BglII fragments were subcloned into the
      pBluescript BamHI site. DNaseI fragments were subcloned
      into the EcoRV site. All subclones were transformed into
      E. coli XL1Blue MRF'. All fragments were then sequenced
      and the sequences where possible were assembled using the
      GCG program GELASSEMBLE."
BASE COUNT      101 a 203 c 175 g 108 t 14 others
ORIGIN

```

```

alignment_scores
  Quality: 124.50      Length: 160
  Ratio: 1.415        Gaps: 5
  Percent Similarity: 55.000      Percent Identity: 28.125

```

alignment_block:

US-09-824-567-2 x B07758/rev ..

Align seg 1/1 to reverse of: B07758 from: 1 to: 601

```

78 IleGluProAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuTh 94
534 ATGCGCCCGGCTTCGGCGAGATACGAGATCTCGGAGACGGCTCAC 485
94 rTyrThrPheLysLeuLysSerAlaPhe...TrpSerAsnGlyAspProL 110
484 CTANACGTTCAAGNTGCGCCGCGGTGAATTCACACGCGCGGAGA 435
110 euThrAlaGluAspPheLysLeuLysGlnValAlaThrGlnGlu 126
434 TGACGCGCGAGGATGTGAATATTCGTCGACCGCGGTGACCTGCGG 385
127 ValSerGlyLysLeuAlaPheAlaLeuAsnProLysAsnValArgly 143
384 ACCCAGTCGCGCGCGCGGCTTTCGCTCGATCAAGGCTTCGATGC 335
143 sIleGlnGluGlyHisLeuSerIleAspHisPheGlyValHisSerProA 160
334 GATGCGCAGCGCTCGCGCAGCAGCTCGAG...GGCGTGACGGTGGTGG 288
160 enGluSerThrLeuValThrLeuGluSerProThrSerHisPheLeu 176
287 ATCCCTCGACGCTCAGGATCGAGCTCTGCGTCCGCGACCCACCTTCCTG 238
177 LysLeuLeuAlaLeuProValPhePheProValHisLysSerGlnArgTh 193
237 CATGTGATGGCGTGAANTTCGCTCGGTGGTGGCGAAGGACCGCTCGA 188
193 rLeuGlnSerLysSerLeu.....ProLysAlaSerGlyAlaPheT 207
187 GGCGCGCGCGCGGANTTCGCGACGCGGTCGCGCAGCGGCGCTTC. 139
207 yrProLysAsnLysGlnLysGlnTrp.....LysLys 218
138 .....AAGCTCGCGAAGTGGACGCTGGCGCGCGCTCGTCTG 103
219 LeuSerLysAsnProHisTyrAsnGln 228
102 TTCGAGAAAGACCGCGACTACTGCGCGAG 73

```

seq_name: gb_gss:A0990866

seq_documentation_block:

```

LOCUS      A0990866                613 bp    DNA    linear    GSS 14-AUG-2000
DEFINITION Rf001703 Photorhabdus luminescens strain W14 M13 library
            Photorhabdus luminescens genomic clone PLG01703, DNA sequence.
ACCESSION  A0990866
VERSION    A0990866.1
KEYWORDS   GSS.
SOURCE     Photorhabdus luminescens.
           Photorhabdus luminescens.
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Photorhabdus.
REFERENCE  1 (bases 1 to 613)
AUTHORS   fFrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
           Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE     A genomic sample sequence of the entomopathogenic bacterium
           Photorhabdus luminescens w14: potential implications for virulence
JOURNAL   Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE   20378633
COMMENT   Contact: french-Constant RH
           Department of Biology and Biochemistry
           University of Bath
           South Building, Bath BA2 7AY, UK
           Tel: (44) 1225 826621
           Fax: (44) 1225 826779
           Email: bssrf@bath.ac.uk
           This is one of 2,122 random reads from the M13 library. For

```

annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res.

Seq primer: M13 Forward

Class: Shotgun.

FEATURES

```

source
  1..613
    /organism="Photorhabdus luminescens"
    /strain="W14"
    /db_xref="taxon:29488"
    /clone_lib="PLG01703"
    /clone="Photorhabdus luminescens strain W14 M13 library"
    /dev_stage="primary phase variant"
    /notes="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

```

BASE COUNT 152 a 139 c 130 g 187 t 5 others
ORIGIN

alignment_scores:

Quality: 122.50 Length: 241
Ratio: 1.021 Gaps: 9
Percent Similarity: 49.793 Percent Identity: 21.992

alignment_block:

US-09-824-567-2 x A0990866/rev ..

Align seg 1/1 to reverse of: A0990866 from: 1 to: 613

```

223 ProHisTyrTyrAsnGlnSerGlnValGluThrLysThrIleThrIleH 239
606 CCAGNATATTACCGCTATCCAGTGAAGATAAATATACCAACAGCG 557
239 sPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysL 256
556 TTATTCCCGGACGCT..... 541
256 euAsnTrpGlnGlyProProTrpGlyGluArgIleProGlnGluThrLeu 272
540 .....TGTATC 535
273 SerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValAlaGlyTh 289
534 ACAAATNAAGAAAGAGA.....TTCNTGATCAGGTTCCG 500
289 rSerTrpLeuThrPhe.....AsnIleAsnLysPheProLeuAsn.... 302
499 GTCAGTCTTACCTTTGTACGTTTATTACGAATTTACCAATAAAACCC 450
303 .....AsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAsp 315
449 GCGGTTTACTTGATGTCAGAGTCCGAGCGCTCAAGTTAAGTCTCGAT 400
316 LysGluAlaLeuValSerThrIlePheLeuGlyArgAlaLys..... 329
399 AGGGATTAATTCGCGAAATTT...ATGGCGGCGGCAATACCGGCT 353
330 .....ThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProG 344
352 TAGGTTTACGCCAACATTTATCGGTGGCGCGGATTTTGTAAAGCCAG 303
344 LuHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeu 360
302 AATGGGCTAATTGGACAGACAGCGCAATTAACGCCGAGAAATG 253
361 PheLysGluAla.....LeuGluLysLeuGlnIleThrAlaLys 373
252 TTAGCTCAGCGCGGTTTCAATCAGCGCAATCCATTGAAGTTTCAACA. 208
373 sAspLeuGluHisLeuAsnLeuLysPheProValSerSerAlaSerS 390
207 .....TTGCTATACAAACCTCTGAGCAATAAATAC 177

```

annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res.

Seq primer: M13 Forward

Class: Shotgun.

FEATURES

```

source
  1..613
    /organism="Photorhabdus luminescens"
    /strain="W14"
    /db_xref="taxon:29488"
    /clone_lib="PLG01703"
    /clone="Photorhabdus luminescens strain W14 M13 library"
    /dev_stage="primary phase variant"
    /notes="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

```

BASE COUNT 152 a 139 c 130 g 187 t 5 others
ORIGIN

alignment_scores:

Quality: 122.50 Length: 241
Ratio: 1.021 Gaps: 9
Percent Similarity: 49.793 Percent Identity: 21.992

alignment_block:

US-09-824-567-2 x A0990866/rev ..

Align seg 1/1 to reverse of: A0990866 from: 1 to: 613

```

223 ProHisTyrTyrAsnGlnSerGlnValGluThrLysThrIleThrIleH 239
606 CCAGNATATTACCGCTATCCAGTGAAGATAAATATACCAACAGCG 557
239 sPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysL 256
556 TTATTCCCGGACGCT..... 541
256 euAsnTrpGlnGlyProProTrpGlyGluArgIleProGlnGluThrLeu 272
540 .....TGTATC 535
273 SerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValAlaGlyTh 289
534 ACAAATNAAGAAAGAGA.....TTCNTGATCAGGTTCCG 500
289 rSerTrpLeuThrPhe.....AsnIleAsnLysPheProLeuAsn.... 302
499 GTCAGTCTTACCTTTGTACGTTTATTACGAATTTACCAATAAAACCC 450
303 .....AsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAsp 315
449 GCGGTTTACTTGATGTCAGAGTCCGAGCGCTCAAGTTAAGTCTCGAT 400
316 LysGluAlaLeuValSerThrIlePheLeuGlyArgAlaLys..... 329
399 AGGGATTAATTCGCGAAATTT...ATGGCGGCGGCAATACCGGCT 353
330 .....ThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProG 344
352 TAGGTTTACGCCAACATTTATCGGTGGCGCGGATTTTGTAAAGCCAG 303
344 LuHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeu 360
302 AATGGGCTAATTGGACAGACAGCGCAATTAACGCCGAGAAATG 253
361 PheLysGluAla.....LeuGluLysLeuGlnIleThrAlaLys 373
252 TTAGCTCAGCGCGGTTTCAATCAGCGCAATCCATTGAAGTTTCAACA. 208
373 sAspLeuGluHisLeuAsnLeuLysPheProValSerSerAlaSerS 390
207 .....TTGCTATACAAACCTCTGAGCAATAAATAC 177

```

```

390 erLeuLeuValGlnLeuLeuArgGluGlnTrpLysGluSerLeuGlyPhe 406
   ::      ::      ::      ::      ::      ::      ::      ::
176 AGCAGGCTATC...GCCGCCGATCCATGCTGGAAAGAAATTTGGGAGC 130
   ::      ::      ::      ::      ::      ::      ::      ::
407 AlalleProIleValGlyLysPheAlaLeuLeuGlnAlaAspLeuSe 423
   ::      ::      ::      ::      ::      ::      ::      ::
129 GATGTGATTTTACAGATTCAGAAATGGAAACCTCTTGAAGCGCGCA 80
   ::      ::      ::      ::      ::      ::      ::      ::
423 rSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAla 440
   ::      ::      ::      ::      ::      ::      ::      ::
79 TCAAGGAATATGAGTGACAGACCACTGGTGTCTGCTATTACAATG 30
   ::      ::      ::      ::      ::      ::      ::      ::
440 sProMetAlaPheLeuThrile 447
   ::      ::      ::      ::      ::      ::      ::      ::
29 AACCTACAGCATTTTAAACATG 7

```

seq_name: gb_gss:AZ302752

```

seq_documentation_block:
LOCUS AZ302752 402 bp DNA linear GSS 06-MAR-2001
DEFINITION GSSBU1690 Brucella abortus random genomic library Brucella
melitensis biovar Abortus genomic clone UU1690, DNA sequence.
ACCESSION AZ302752
VERSION AZ302752.1 GI:10128963
KEYWORDS GSS.
SOURCE Brucella melitensis biovar Abortus.
ORGANISM Brucella melitensis biovar Abortus
Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
Brucellaceae; Brucella.

```

```

REFERENCE
AUTHORS Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,
Faccio,P., Diaz,G., Lanzavecchia,S., Aguero,F., Frasch,A.C.C.,
Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.
JOURNAL Gene discovery through genomic sequencing of Brucella abortus
COMMENT Infect. Immun. 69 (2), 865-868 (2001)
Contact: Siv Andersson
Small Genomes Sequencing Group
Department of Molecular Evolution, Uppsala University
Norbyvagen 18C, S-752 36, Uppsala, Sweden
Tel: 46-18-471-4379
Fax: 46-18-471-6404
Email: Siv.Andersson@ebc.uu.se
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Class: shotgun.
Location/Qualifiers
SOURCE 1. 402
/organism="Brucella melitensis biovar Abortus"
/strain="2308"
/db_xref="taxon:235"
/clone="UU1690"
/clone.lib="Brucella abortus random genomic library"
/note="Vector: modified M13"

```

```

BASE COUNT 105 a 110 c 98 g 89 t
ORIGIN
alignment_scores:
Quality: 118.00 Length: 130
Ratio: 1.595 Gaps: 6
Percent Similarity: 56.923 Percent Identity: 27.692

```

```

alignment_block:
US-09-824-567-2 x AZ302752 ..
Align seq 1/1 to: AZ302752 from: 1 to: 402

```

```

220 SerLysAsnProHisTyrTyrrAsnGlnSerGlnValGluThrLysThrII 236
   ::      ::      ::      ::      ::      ::      ::      ::
8 ACCCGCATGATAGCTACTGGGGGCCAAGCGCTGCT...TATGAGCAGAT 54

```

```

236 eThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnG 253
   ::      ::      ::      ::      ::      ::      ::      ::
55 AACTCTCAAGGTATACCCGATCCGAATACACGGCGGATTCGCTCGAGG 104
   ::      ::      ::      ::      ::      ::      ::      ::
253 IngLysLeuAsn.....TrpGlnGly.....ProPro 262
   ::      ::      ::      ::      ::      ::      ::      ::
105 CTGGCGAAATAGACCTGATCTATGSAACGAGCGCGGATTCGCCCGA 154
   ::      ::      ::      ::      ::      ::      ::      ::
263 TrpGlyGluArgIleProGlnGluThrLeuSerAsnLeuGlnSerLysgl 279
   ::      ::      ::      ::      ::      ::      ::      ::
155 TACTTTCGAGCGCTCCAGAA.....AATGG 180
   ::      ::      ::      ::      ::      ::      ::      ::
279 yHisLeu.....HisSerPhe AspValAlaGlyThrSerTrpLeuThr 293
   ::      ::      ::      ::      ::      ::      ::      ::
181 GCATCTATATACCGAGCTTCGGAGCGCTTAGAAACGTCGCTTTTGGCG 230
   ::      ::      ::      ::      ::      ::      ::      ::
294 PheAsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGluAlaLe 310
   ::      ::      ::      ::      ::      ::      ::      ::
231 CTCACACCAATCACGCGCAACAAAGAACCTCGCTCGCGCAAGCCAT 280
   ::      ::      ::      ::      ::      ::      ::      ::
310 uAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyA 327
   ::      ::      ::      ::      ::      ::      ::      ::
281 CAATCATGCAGTCGATAGGACAGCATGATTCGCCACCGTCTCTACGGGA 330
   ::      ::      ::      ::      ::      ::      ::      ::
327 rgAlaLysThrAlaAspHisLeuLeuProThrAsnIle 339
   ::      ::      ::      ::      ::      ::      ::      ::
331 CACAGAAGCGCTGCAGATACCTGTTGCCGATATGTT 368

```

seq_name: gb_gss:AQ012177

seq_documentation_block:

```

LOCUS AQ012177 713 bp DNA linear GSS 04-JUN-1998
DEFINITION 271IC073112697 Cosmid library of chromosome II Rhodobacter
sphaeroides genomic clone 271IC073112697, DNA sequence.
ACCESSION AQ012177
VERSION AQ012177.1 GI:3177000
KEYWORDS GSS.
SOURCE Rhodobacter sphaeroides.
ORGANISM Rhodobacter sphaeroides
Bacteria: Proteobacteria; alpha subdivision; Rhodobacter group;
Rhodobacter.

```

REFERENCE 1 (bases 1 to 713)

```

AUTHORS Choudhary,M., Mackenzie,C., Mouncey,N., Weinstein,G.M. and Kaplan
,S.
TITLE RSGDB, the Rhodobacter sphaeroides Genome Database
JOURNAL Unpublished (1998)
COMMENT Contact: Choudhary, M.
Department of Microbiology and Molecular Genetics
University of Texas Medical School
6431 Fannin Street, Houston, TX 77030, USA
Tel: 713 500 5437
Fax: 713 500 5499
Email: madhu@utmsi.med.utmc.edu
Seq primer: pbluescript T3
Class: shotgun.
Location/Qualifiers
SOURCE 1. 713
/organism="Rhodobacter sphaeroides"
/strain="2.4.1r"
/db_xref="taxon:1063"
/clone="271IC073112697"
/clone.lib="Cosmid library of chromosome II"
/lab_host="E. coli S17-1"
/note="Vector: pLA2917"

```

```

BASE COUNT 132 a 250 c 216 g 115 t
ORIGIN
alignment_scores:
Quality: 114.50 Length: 172
Ratio: 1.180 Gaps: 5
Percent Similarity: 56.395 Percent Identity: 23.256

```


enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and DnaSeI) and restriction fragments subcloned into the respective multiple cloning site sites of pBluescript SK (-). Note BglII fragments were subcloned into the pBluescript BamHI site. DnaSeI fragments were subcloned into the EcoRV site. All subclones were transformed into *E. coli* XLBlue MRF'. All fragments were then sequenced and the sequences where possible were assembled using the GCG program GELASSEMBLE.

BASE COUNT 83 a 137 c 171 g 88 t 19 others
ORIGIN

alignment_scores:
Quality: 104.00 Length: 145
Ratio: 1.268 Gaps: 7
Percent Similarity: 56.552 Percent Identity: 30.345

alignment_block:
US-09-824-567-2 x B07745/rev ..

Align seg 1/1 to reverse of: B07745 from: 1 to: 498

```

93  LeuThrTyrThrPheLysLeuLysSerAlaPheTyrSerAsnGlyAspPr 109
|||||
362  CTCACCTACAGTTCCATCTGCGACGGGAC.TTCTCCGCGGCTCGCC 314
|||||
109  oLeuThrAlaGluAspPheLeuGluSerTyrLysGlnValAlaThrGln 126
|||||
313  CATCACCCGAGGAC..... 298
|||||
126  luValSerGlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArg 142
|||||
297  .....GCGGCTTCCAGCTG.....CTG 280
|||||
143  LysIleGlnGluGlyHisLeuSerIle.....AspHisPheGlyValHI 157
|||||
279  CGCATCCGCGACAGCAACTGCTGCTGTCGGACAGCTATCGGTGAT 230
|||||
157  sSerProAsnGluSer.....ThrLeuValValThrLeuGluS 170
|||||
229  CGAACTNCCGAGCGAGCCAGCCGACGACCTGCTGCTCAAGCTGAAGA 180
|||||
170  erProThrSerHisPheLeuLysLeuAlaLeuProValPhePhePro 186
|||||
179  CCCCTCGCGCCCTTCTGTCACCATCGGATGCCGCGCTCGATC 130
|||||
187  ValHisLysSer.....GlnArgThrLeuGlnSerLy 197
|||||
129  CTGTCGAGGCGGCGTCAAGCCATGGGCGAGGAGGCTATCCGAGAA 80
|||||
197  sSerLeuProIleAlaSerGlyAlaPheTyr.ProLysAsnIleLysGln 213
|||||
79  G.....CCGCGTGGCTCGGCGCTTTACCGTCCAGGAATGCCGCC 36
|||||
214  LysGlnThrPrleLysLeuSerLysAsnProHis 224
|||||
35  GCGCACCCAGCTATCTCTAAANANAACCCGAAT 3
|||||
seq_name: gb_gss:TA380G06P

```

seq_documentation_block:
LOCUS TA380G06P 519 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 380G06, forward sequence,
genomic survey sequence.

ACCESSION AL497735
VERSION AL497735.1 GI:11873457

KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei

REFERENCE 1 Trypanosoma.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 519)

AUTHORS

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Meiville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE

Direct Submission

JOURNAL

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 Gurat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

Location/Qualifiers

source

1..519

/organism="Trypanosoma brucei"

/strain="TREU927"

/db_xref="taxon:5691"

/clone="380G06"

BASE COUNT 127 a 115 c 128 g 149 t

ORIGIN

alignment_scores:

Quality: 103.50 Length: 143

Ratio: 1.344 Gaps: 4

Percent Similarity: 53.846 Percent Identity: 24.476

alignment_block:

US-09-824-567-2 x TA380G06P ..

Align seg 1/1 to: TA380G06P from: 1 to: 519

```

250  LeuPheAsnGlnGly...LysLeuAsnTyrGlnGlyProTyrGlyG 265
|||||
110  TTAGAAATAGGGCGCACAAATGTTATGGAGCACCCGCCACAGTCGA 159
|||||
265  uArgIleProGln.....G 270
|||||
160  AAGGTTACCGCAGGTTTCGACGACGATGGAGGCTGCCAGCACCA 209
|||||
270  luThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspVal 286
|||||
210  GTACACTCGAGCACTTCAGAGTTCTTCTGTTCTTCACACATTTGATAC 259
|||||
287  AlaGlyThrSerTyrLeuThrPheAsnIleAsnLysPheProLeuAsnAs 303
|||||
260  TCAGGGCCATTCTCAGACCTTTCTCCGTCG..... 289
|||||
303  nMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeu 320
|||||
290  .....CTCCTCTCATGTCGCAACATTTTGGTGGATTCAGCGTAAGTC 332
|||||
320  alSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeuPro 336
|||||
333  AAAACACATGTCATATGCGGCAAGCGGTGAATTTGACGTTTGGCG... 379
|||||
337  ThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMetAlaGlnAr 353
|||||
380  .....CGTGAGGTGAAGACTTCGCGGCACACTTGTCTCTCA 417
|||||
353  gGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluGlnI 370
|||||
418  GCAGAGTTACTTTTCAGGATCTGGAGAGTCGTAGATTACACACTCAGTGG 467
|||||
370  leThrAlaLysAspLeuGluHisLeuAsn 379

```

468 CTTGACGAGGAGGAGCGCAAGGAATCAAC 496
 seq_name: gb_gss:BH398784

seq_documentation_block:
 LOCUS BH398784 554 bp DNA linear GSS 11-DEC-2001

DEFINITION AG-ND-150B19, TR ND-TAM Anopheles gambiae genomic clone AG-ND-150B19
 , DNA sequence.

ACCESSION

VERSION BH398784.1 GI:17345000

KEYWORDS

SOURCE

ORGANISM

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 554)

Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.

Direct Submission of BAC-end sequences from Anopheles gambiae

Unpublished (2001)

Other_GSSs: AG-ND-150B19, TF

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.

Seq primer: M13 Rev

Class: BAC ends.

FEATURES

Source

Location/Qualifiers

1..554

/organism="Anopheles gambiae"

/strain="PEST"

/db_xref="taxon:7165"

/clone="AG-ND-150B19"

/clone_lib="ND-TAM"

/note="Vector: pECBAC1; Site_1: HindIII"

155 a 132 c 92 g 175 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 103.50 Length: 176

Ratio: 1.015 Gaps: 5

Percent similarity: 57.955 Percent Identity: 20.455

alignment_block:

US-09-824-567-2 x BH398784/rev ..

Align seg 1/1 to reverse of: BH398784 from: 1 to: 554

290 SerTrpLeuThrPheAsn.....IleAsnLysPheProLe 301

550 ACCTATACCATGTTCAATATGAGGATCCGATGTGGGGGTATAGTCT 501

301 uAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluA 318

500 AGACAAAATAGCGTACCGCGTCCATTACCTTGGCATATACCAACAGG 451

318 laLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeu 334

450 AGCGGATTAAAGCACTTATAAAGGTCAAGCGGTAGAGCTGAAATCTTG 401

335 LeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMetAl 351

400 GTGCCAGATGGGGTGCAGGGCACAATGCTCAATATAGAGTAGTGTGTC 351

351 agInArgGlnAlaTyrAlaLysLysLeu.....PheLysG 363

350 CTATACCCGTTATTGGCCCAATAATTACTGGATCGATTGGTTATAAGA 301

363 luAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGluHisLeuAsn 379

300 AAGCAAAAGATGGTTATCGA...ACTTTACCAACAGCGGCAGCCATTACC 254

380 LeuIlePheProValSerSerAlaSerLeuValGlnLeuI 396

253 CTGAATATATATCTGAAATAGCTCAAGCTCAGTCATTCATCTCT.... 209

396 eArgGluGlnTrpLysGluSerLeu.....GlyPheAlaIleProI 410

208GAGTTATGGAAGAAAACCTTAGATGCTATTGGGGTCCGTGTGACT 163

410 leValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerGlyAsn 426

162 TTAGGTCAGTAATTTTGGCGATAATTTAAAGAGAGCCATGCAATGTAAA 113

427 PheSerLeuAlaThrGlyTrpPheAlaAspPheAlaAspProMetAl 443

112 CATATGATTTGGGTGGCGCATGGATGCTGATTTTCTCTGAAGGAGAAAA 63

443 aPheLeuThrIlePheAlaTyrProSer 452

62 TTTTGCTCAGTTGCTTTATGTCGTCGCAAC 35

seq_name: gb_gss:TA358F03P

seq_documentation_block:

LOCUS TA358F03P 558 bp DNA linear GSS 13-DEC-2000

DEFINITION T. brucei sheared genomic DNA clone 358f03, forward sequence,

genomic survey sequence.

ACCESSION AL494115

VERSION AL494115.1 GI:11870744

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 558)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,

Chillingworth, C., Ormond, B., Harris, B., El-Sayed, N., Hou, L.,

Melville, S.E., Rajandream, M.A. and Barrell, B.G.

Direct Submission

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nhles@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsaved@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at <http://www.sanger.ac.uk/projects/T-brucei/>.

FEATURES

source

1..558

/organism="Trypanosoma brucei"

/strain="TREU927"

/db_xref="taxon:5691"

/clone="358f03"

AUTHORS
TITLE
JOURNAL
COMMENT

Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.
 Direct Submission of BAC-end sequences from *Anopheles gambiae*
 Unpublished (2001)
 Other_GSSs: AG-ND-174G16.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from *A. gambiae* PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.

Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 Location/Qualifiers

1..583
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-174G16"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 161 a 133 c 99 g 190 t

alignment_scores:

Quality: 102.50 Length: 176
 Ratio: 1.005 Gaps: 5
 Percent Similarity: 57.955 Percent Identity: 20.455

alignment_block:

US-09-824-567-2 x BH375296/rev ..

Align seg 1/1 to reverse of: BH375296 from: 1 to: 583

```

290 SerTrpLeuThrPheAsn.....IleAsnLysPheProLeu 301
      ::::: |||||
555 ACCTATACCATGTCATATCAAGGATCCGATTGTGGGGGGTATAGTCT 506
      ||:::|||||
301 uAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluA 318
      ||:::|||||
505 AGACAAATAGCGGTACGCGCGCCATTACCTTGGCATATACCAACAGG 456
      ||:::|||||
318 laLeuValSerThrIlePheLeuGlyArgAlaLeuThrAlaAspHisLeu 334
      ::::: |||||
455 AAGCGATTAAAGCACTTTATAAGGTCAAGCGTTAGAGCTGAATGTTG 406
      ||:::|||||
335 LeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMetAl 351
      ||:::|||||
405 GTGCCAGATGGGTGCAAGGGCACAATGCTCAATATAGAGTAGTGTTC 356
      ||:::|||||
351 aGlnArgGlnAlaTyrAlaLysLysLeu.....PheLysG 363
      ||:::|||||
355 CTATAACCCGGTATTGGCCCAATAAATTACTGGATCGATTGGTTATAAGA 306
      ||:::|||||
363 luAlaLeuGluLeuGlnIleThrAlaLysAspLeuGluHisLeuAsn 379
      ||:::|||||
305 AGGCAAGAGATGGTTATCGA...ACTTTACCAACAGGGCAGCCATTACC 259
      ||:::|||||
380 LeuIlePheProValSerSerAlaSerSerLeuLeuValGlnLeuIrl 396
      ||:::|||||
258 CTGAATATATATAGTAATAGCTCAGCTCAGCTCAGTTCATTCTCT... 214
      ||:::|||||
396 eaArgGluGlnTrpLysGluSerLeu.....GlyPheAlaIleProI 410
      ||:::|||||

```

```

213 ...GAGTTATCGAAGAAACCTAGATGCTATTGGGGTAATCTTCACT 168
      ||| |||||
410 leValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsn 426
      :::::|||||
167 TTAAGTCAAGTAATTTTGGCGGATAATTTAAAGAAAGCCATGCAATGTAAA 118
      ||| |||||
427 PheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAl 443
      ||| |||||
117 CATATGATTGGGGTGGGGCATGATGCTGATTCTTCTGAAGGAGAAAA 68
      ||| |||||
443 aPheLeuThrIlePheAlaTyrProSer 452
      ||| |||||
67 TTTTCTCAGTTGCTTTATGTCGGAAC 40

```

OM of: US-09-824-567-2 to: GenEmbl.* out_format : pfs
 Date: Jul 26, 2002 5:53 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed_p2n.model -DEV=xlp
 -O=/cgn2_1/USPTO_spool/US09824567/runat_23072002_141350_13449/app_query.fasta_1.596
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -DELEXI=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEADSIZE=500
 -MINLEN=0 -MAXLEN=200000000 -USER=US09824567 -CCGNL_1.7895
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPAY -WAIT -THREADS=1

Search information block:

Query: US-09-824-567-2
 Query length: 532
 Database: GenEmbl.*
 Database sequences: 1797656
 Database length: 187333701
 Search time (sec): 2484.090000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb.pat:AX349501	+ 2739.00	4065.71	3.7e-218	1599	AX349501 Sequence 24 from Patent
gb.pat:AX268341	+ 2739.00	4065.56	4.3e-218	1799	AX268341 Sequence 1 from Patent
gb.ba:AE001606	+ 2739.00	4047.44	4.4e-217	11648	AE001606 Chlamydia pneumoniae
gb.ba:AE002216	+ 2739.00	4047.34	4.5e-217	11764	AE002216 Chlamydia pneumoniae
gb.ba:AP002545	+ 2739.00	4015.92	2.5e-215	296650	AP002545 Chlamydia pneumoniae
gb.ba:AE001293	+ 1345.00	1998.04	6.3e-103	11944	AE001293 Chlamydia pneumoniae
gb.ba:AE002315	+ 1345.00	1969.54	2.4e-101	12173	AE002315 Chlamydia pneumoniae
gb.pat:AX349605	+ 1160.50	1714.37	4.0e-87	1584	AX349605 Sequence 128 from Patent
gb.ba:AE002215	+ 1160.50	1693.96	5.4e-86	12980	AE002215 Chlamydia pneumoniae
gb.pat:AX349505	+ 755.50	1110.79	1.7e-53	1587	AX349505 Sequence 28 from Patent
gb.pat:AX300389	+ 748.00	1109.63	1.9e-53	1787	AX300389 Sequence 1 from Patent
gb.ba:AE002313	+ 748.00	1090.47	8.1e-52	11402	AE002313 Chlamydia pneumoniae
gb.ba:AE001291	+ 685.50	987.83	1.2e-46	10827	AE001291 Chlamydia pneumoniae
gb.ba:BA000007	+ 617.50	891.75	2.6e-41	6300	M57689 Bacillus subtilis sp008
gb.ba:BA000007	+ 617.50	891.75	2.6e-41	6300	AX027646 Sequence 1 from Patent
gb.ba:BS000007	+ 617.50	857.41	2.1e-39	216750	AX027646 Sequence 1 from Patent
gb.ba:BS000007	+ 614.50	888.14	4.2e-41	5760	X56347 B. subtilis opp operon
gb.ba:AE007859	+ 536.50	763.20	3.8e-34	14123	AE007859 Clostridium acetobutylicum
gb.ba:BA0250013	+ 509.50	744.37	4.4e-33	1587	AJ250013 Borrelia garinii p58
gb.ba:BA0250012	+ 506.00	738.96	8.5e-33	1587	AJ250012 Borrelia afzelii p58
gb.ba:AE001139	+ 505.50	724.78	5.2e-32	6336	AF043071 Borrelia burgdorferi
gb.ba:AE001139	+ 505.50	719.54	1.0e-31	10867	AF001139 Borrelia burgdorferi
gb.ba:AF000366	+ 505.50	718.41	1.2e-31	12206	AF000366 Borrelia burgdorferi
gb.ba:BSA02571	+ 489.50	679.86	1.7e-29	55593	AJ002571 Bacillus subtilis 168
gb.ba:BSA02571	+ 483.50	692.85	3.1e-30	5798	X56678 B. subtilis acta operon
gb.pat:AX349503	+ 482.50	705.81	6.0e-31	1308	AX349503 Sequence 26 from Patent
gb.ba:RNE296269	+ 479.50	693.26	3.0e-30	3007	AJ296269 Rhizobium meliloti op
gb.ba:RNE296269	+ 479.50	648.40	9.4e-28	306050	AL603645 Rhizobium meliloti
gb.ba:AF247635	+ 476.00	678.18	2.1e-29	8314	AF247635 Lactococcus lactis pe
gb.ba:AE009355	+ 474.50	671.93	4.6e-29	12571	AE009355 Agrobacterium tumefaci
gb.ba:AE008258	+ 474.50	671.93	4.6e-29	12571	AE008258 Agrobacterium tumefaci
gb.ba:AE006271	+ 474.00	641.84	2.2e-27	258650	AL596171 Listeria innocua Cl
gb.ba:AE006271	+ 472.00	669.33	6.4e-29	11198	AF006271 Lactococcus lactis
gb.ba:AE000948	+ 471.50	679.83	1.7e-29	3515	AF000948 Borrelia burgdorferi
gb.ba:AE000792	+ 471.50	660.22	2.1e-28	26498	AE000792 Borrelia burgdorferi
gb.ba:AE003793	+ 470.50	673.09	4.0e-29	6035	AF103793 Listeria monocytogenes
gb.ba:AE001982	+ 470.50	635.34	5.0e-27	295050	AL591982 Listeria monocytogenes
gb.ba:AE003793	+ 467.50	679.97	1.6e-29	1873	S75873 60 kDa antigen [Borrelia]
gb.ba:AE003793	+ 467.50	679.97	1.6e-29	1873	S75873 60 kDa antigen [Borrelia]
gb.ba:AE003793	+ 466.50	599.25	5.1e-25	30050	AL596173 Listeria innocua Cl
gb.ba:AE003793	+ 446.00	632.31	7.4e-27	9364	AF305387 Bacillus thuringiensis

gb.ba:AF465631 + 441.50 626.24 1.6e-26 8771 ! AF465631 Oenococcus oeni le
 gb.ba:AE002565 - 441.00 611.58 1.1e-25 36804 ! AE002565 Enterococcus faec
 gb.ba:AF188935 - 440.50 601.51 3.8e-25 96231 ! AF188935 Bacillus anthraci
 gb.ba:ECUR8242 + 438.50 636.69 4.2e-27 1886 ! U88242 Escherichia coli per
 gb.ba:AE000231 + 438.50 618.11 4.6e-26 12790 ! AE000231 Escherichia coli
 seq_name: gb_pat:AX349501
 seq_documentation_block:
 LOCUS AX349501 1599 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 24 from Patent WO0202606.
 ACCESSION AX349501
 VERSION AX349501.1 GI:18615357
 KEYWORDS
 SOURCE Chlamydia pneumoniae.
 ORGANISM Chlamydia pneumoniae.
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 1 (sites)
 AUTHORS Ratti, G. and Grandi, G.
 TITLE Immunisation against Chlamydia pneumoniae
 JOURNAL Patent: WO 0202606-A 24 10-JAN-2002;
 Chiron S.p.A. (IT)
 FEATURES
 source location/Qualifiers
 1.1599
 /organism="Chlamydia pneumoniae"
 /db_xref="taxon:83558"
 BASE COUNT 509 a 389 c 261 g 440 t
 ORIGIN

alignment_scores:
 Quality: 2739.00 Length: 532
 Ratio: 5.148 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-824-567-2 x AX349501
 Align seg 1/1 to: AX349501 from: 1 to: 1599
 1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
 1 ATCCGCAAGATATCATGTGGGAATCTGTATCACCATTCTCTAGCCTC 50
 17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34
 51 CGTAGCTCTCCAGAGCTGCAGAGGTCAGTCTCTCTATCATCTCGG 100
 34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
 101 GAGAACTCGTATTATAAAGAGATGAACCCCGTCTTTAGATCCAGA 150
 51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluG 67
 151 CAAGTCGCGACTCTTTTTCAGAAATCAGCCTGTGCAACATATCTATGAGG 200
 67 yLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
 201 ATTAGTTCAAGAAATATATCTTTTCAGAAATATAGAGCTCTCTTCGAC 250
 84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
 251 AAGACTACTCTCTTCTCGAGCGGACTCACATATACATTTTAAACTGAAA 300
 101 SerAlaPheTyrSerAsnGlyAspProLeuThrAlaGluAspPheLeuG 117
 301 TCAGCTTTTGGAGTATATGCGACCCCTTAAACAGCTGAAGACTTTATAGA 350
 117 uSerTyrLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPhea 134
 351 ATCTTGGAAACAACTACTACTCAAGAGCTCAGGAATCTATGCTTTT 400
 134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150

```

|||||
401 CCTTGATCCCAATTAATAATGTCGAAAGATCCAGAGGACACCTCTCC 450
|||||
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 467
|||||
451 ATAGACCAATTTTGAGTGCACCTCTCTATGAATCATACACTTGTGTGTTAC 500
|||||
167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValp 184
|||||
501 COTGGAATCCCAACCTCGCAATTTCTTAAACTTTTAGCTTCTCCAGTCT 550
|||||
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
|||||
551 TTTTCCCGGTTCAATAATCTCAAGAACCTGCAATCCAAATCTCTACCT 600
|||||
201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnIleThrL 217
|||||
601 ATAGCAAGCGAGCTTCTTATCTCTTAAATAATATCAACAAACAATGGAT 650
|||||
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
|||||
651 AAAACTCTCAAAAACCCCTCACTACTATATAATCAAAAGTCAGGTGCAAACTA 700
|||||
234 ysThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
|||||
701 AAACGATTACGATTCACCTTATCCCGATGCAAAACACAGCAGCAAACTA 750
|||||
251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgI 267
|||||
751 TTTAATCAGGGAATACTCAATTTGCAAGGACCTCTTGGGAGAACGAT 800
|||||
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
|||||
801 TCCCTCAAGAAACCTTATCAATTTACAGTCTAAGGGCACTTACACTCTT 850
|||||
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
|||||
851 TTGATGTCGAGAGAACCTCATGGCTCACCTTCAATATCAATAAATCCCC 900
|||||
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysgl 317
|||||
901 CTCACAAATATGAAGCTTAGAGAGCCTTAGCATCAGCCTTAGAAGA 950
|||||
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
|||||
951 AGCTTTGTCTCAACTATATTTCTTAGCCGCTGCAAAACTGCCGATCATC 1000
|||||
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
|||||
1001 TCCTACCTACAATATTCATAGCTATCCCAACATCAAAACAAGAGATG 1050
|||||
351 AlaGlnArgGluAlaTyrAlaLysLysLeuPheLysGluAlaLeuGlucl 367
|||||
1051 GCACAACGGCAAGCTTAGCGTAAAAAACTCTTAAAGAGCTTTAGAAGA 1100
|||||
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePhePro 384
|||||
1101 ACTCCAAATCACTGTAAGAATCTCGAACAATCTTAATCTTATCTTCCGG 1150
|||||
384 alSerSerAlaSerSerLeuLeuValGlnLeuIleargGluGlnTrp 400
|||||
1151 TTTCTCTGTCACCAAGTTCTTTACTAGTCCAACTTATAGGAGAACAGTGG 1200
|||||
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
|||||
1201 AAAGAAAGTTTAGGCTTCGCTATCCCTATTGTCGGAAGGAATTTGCTCT 1250
|||||
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
|||||
1251 TCTCCAGCAGACCTATCTTCTAGGAACTCTCTTTAGCTACAGGAGAT 1300
|||||
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
|||||

```

```

1301 GGTTCGAGACTTTGCTGATCCTATGCAATTTCTAACGATCTTTCTCTTAT 1350
|||||
451 ProSerGlyValProTyrAlaIleAsnHisLysAspPheLeuGluIle 467
|||||
1351 CCATCAGGAGTTCTCTCTTATGCAATCAACCATAGGACTTCTCTAGAAAT 1400
|||||
467 eLeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuV 484
|||||
1401 TCTACAAAACATAGAACAGAGCAAGATCACCAAAAACGCTCGGAATTAG 1450
|||||
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleleuProIle 500
|||||
1451 TGTTCGCAAGCTTCTCTTACCTAGAGACCTTTCATATTTAGGCCGATC 1500
|||||
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuGl 517
|||||
1501 TACCACGACGCAATTTCAATTTGCTATGATGAATAAAACATTTCTAATCTAG 1550
|||||
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
|||||
1551 AGTCTCACCACAGAGGTTGTGGACTTCGCTTATGCTAAGGAAAT 1596
|||||
seq_name: gb_pat:AX268341
seq documentation_block:
LOCUS AX268341 1799 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174863.
ACCESSION AX268341
VERSION AX268341.1 GI:16541562
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae.
REFERENCE
1 (sites)
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE Chlamydia antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0174863-A 1 11-OCT-2001,
Aventis Pasteur Limited (CA)
FEATURES
source
location/Qualifiers
1..1799
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83558"
101..1699
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD10302.1"
/db_xref="GI:16541563"
/translacion="MBKISVGICITILLISLVVLOGCKESSHSTSRGELAINIDPE
RSLDPQVRLLSISLVKHYEGLVQENNLGNIEPALAEDYSLSDGLTYTFLKLSA
FWSNGDPLTAEDEFIESKQVATQEVSGIYAFALNPKNVRKIQEHLSDHFGVHSPN
ESTLVLTESPTSHFLKLLALEVFPFVHKQSRTLQSKSLPIASGAFYKNIKQKOWIK
LSKNPHYNQSVETKTIITHTIPDANAAKLNQKLNQWPPMGERIPQETLSNLO
SKGHLHSDVAGT SWLTFNKFFPLNNMKLREALSALDKALYSTIFLGRAKTADHL
LPTNHSYPEHQKQEAQYAKKLPKEALELQITAKDLEHLNLIFFVSSSSALL
VOLIREQKESLGFALPIVQKEFALLQADLSGNFSIATGGWFAFDADPMAFLTIFAY
PSGVPPIAINHDFLEILQEQDQHQKSELVSQASLYLETFHIEPIYHDAFQFA
MNKLSNLGVSPTGVDFRYAKEN"
BASE COUNT 560 a 439 c 294 g 506 t
ORIGIN

```

```

alignment_scores:
Quality: 2739.00 Length: 532
Ratio: 5.148 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-824-567-2 x AX268341 ..
Align seg 1/1 to: AX268341 from: 1 to: 1799

```

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17


```

101 ATGCCCAAGATATCAGTGGAAATCTGTATCACCATTCTCCTTAGCCTCTC 150
17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34
151 CGTAGCTCTCCAGGCTGCAAGGATCCAGTCACTCTCTACATCTCGGG 200
34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
201 GAGAACTCGCTATTATATATAAGAGATGAACCCCGTTCTTTAGATCCAAGA 250
51 GlnValArgLeuSerGluIleSerLeuValLysHisIleThrGlu 67
251 CAAGTGCACCTCTCTTCAGAAATCAGCCTTGCAACATATATCATGAGG 300
67 yLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAla 84
301 ATTAGTTCAAGAAATAATCTTTTCAGGAATATAGAGCCTGCTCTGACAG 350
84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
351 AAGACTACTCTCTCTCTCGGAGGACTCACTTATACCTTTTAACTGAAA 400
101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117
401 TCAGCTTTTGGAGTAATGCGACCCCTTACAGCTGAAGACTTATAGA 450
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
451 ATCTTGGAAACAAGTAGTACTCAAGAAGTCTCAGGAATCTATGCTTTG 500
134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
501 CCTTGAAATCCAAATTAATAATGACGAAAGATCCAAAGAGGACACCTCTCC 550
151 ileAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
551 ATAGACCAATTTGGAGTGACCTCTCTTAATGAATCTACACTTGTGTAC 600
167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184
601 CCTGGAATCCCAACCTCGCATCTCTTAAACCTTTAGCTCTTCCAGTCT 650
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
651 TTTTCCCGCTGCATAAATCTCAAGACCCCTGCANTCCAAATCTCTACCT 700
201 lleAlaSerGlyIlePheTyrProLysAsnIleLysGlnLysGlnTrpI 217
701 ATAGCAAGGAGGAGCTTTCTATCTCTAAAAATATCAAAACAAACATGAT 750
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
751 AAACCTCTCAAAAACCTCTACTACTATAATCAAACTCAGGTGGAACATA 800
234 ysthrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
801 AAACGATTACGATTCACTTCTATCCCATGCAACACAGCAGCAAACTA 850
251 pheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgI 267
851 TTTAATCAGGAAAACTCAATTTGCAAGGACCTCTCTTGGGAGAACCAT 900
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
901 TCTCTAAGAAACCTTATCCAAATTTACAGTCTAAGGGCAGCTTACACTT 950
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
951 TTGATGTCGAGAACTCATGCTCACCTTCAATATCAATAAATATCCCC 1000
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317

```

```

1001 CTCACCAATATGAAGCTTAGAGAGCCCTTAGCATCAGCCTTAGATAAGA 1050
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
1051 AGCTCTGTGCTCAACTATATCTTAGGCCGTGCAAAACTGCCGATCATC 1100
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
1101 TCTACTACTACAATATTCATAGCTATCCCGAATCAAAACAAAGAGATG 1150
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGlu 367
1151 GCACAACGCCAAGCTTAGCTTAAAAACTCTTTAAAGAGCTTTAGAGAGA 1200
367 uLeuGlnIleThrAlaLysAspLeuGluHisIleAsnLeuIlePhePro 384
1201 ACTCCAAATCACTGCTAAAGATCTCGAATCTTAATCTTATCTTCTCCG 1250
384 alSerSerSerAlaSerSerLeuValGlnLeuIleArgGluGlnTrp 400
1251 TTCTCTCGTCAGCAAGTTCTTTACTAGTCCAACTTATACGAGAACAGTGG 1300
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
1301 AAAGAAAGTTTAGGGTTCCGCTATCCCTATTGTCGAAAGGAATTTGCTCT 1350
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
1351 TCCTCAGCAGACCTTCTCTCAGGAACTTCTTTAGCTACAGGAGAT 1400
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
1401 GGTTCGACAGACTTTGCTGATCCTATGGCATTTCTAACGATCTTTGCTTAT 1450
451 ProSerGlyValProTyrAlaIleAsnHisLysAspPheLeuGluI 467
1451 CCATCAGAGATTCTCTTATGCAATCAACCATAGGACTTCTCTAGAAAT 1500
467 eLeuGlnAsnIleGluGlnGlnAspHisGlnLysArgSerGluLeuV 484
1501 TCTACAAACATAGAACACAGCAGATCCCAAAACGCTCGGAATTAG 1550
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
1551 TGTCGCAAGCTTCTCTTACCTAGAGACCTTCTATTTATTTAGCCGATC 1600
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517
1601 TACCACGACGATTTCAATTTGCTATGATAAAAAAATCTTTCTAATCTAGG 1650
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
1651 AGTCTCACCAACAGGAGTTGTGGACTTCGTTATGCTTAGGAAAAAT 1696

```

seq_name: gb_ba:AE001606

seq_documentation_block:

LOCUS AE001606 11648 bp DNA linear BCT 01-DEC-2000
DEFINITION Chlamydia pneumoniae section 22 of 103 of the complete genome.

ACCESSION AE001606 AE001363

VERSION AE001606.1 GI:4376464

KEYWORDS

SOURCE

ORGANISM

Chlamydia pneumoniae CWL029.

Chlamydia pneumoniae CWL029.

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 11648)

AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,

Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.

TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis

JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

MEDLINE 99206606

PUBMED 10192388

REFERENCE 2 (bases 1 to 11648)

AUTHORS

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.

TITLE
JOURNAL

Direct Submission
Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

FEATURES

Source

1. 11648
/organism="Chlamydia pneumoniae CWL029"

/strain="CWL029"

/db_xref="taxon:115713"

73. 1107

/gene="gcp_2"

/note="Cp0194"

73. 1107

/gene="gcp_2"

/codon_start=1

/transl_table=11

/product="O-sialoglycoprotein Endopeptidase"

/protein_id="AAD18347.1"

/db_xref="GI:4376465"

/translation="MLPLGLESCDETCATVNEKQILANIIASODIHASYGGVPE
LASRAHLHIFPVINKALQOANLIEDMDLIAVQTQGLIGSLSGVGRKGIAIGAK
KSLIGVNEVHAHYAAYMAQNVQPALGLVYSGAHTAFAFFIENPTSYKLGKTRDDA
IGETDKVGRGLGYPAGPLIEKALGSEDSYSPSPAKVPNYSFSGIAIVLYA
IKGNSSPSAPISLEKORDIASFOKACTIIAQKLETTIIEFSCRSILIGGVA
INEPRAIQTACNLPPVTPPAKLCSDNAIAGLGGENFQKNSIPEIRICARYONE
SVSPRSASP"

1053. 2651

/gene="oppA_1"

/note="Cp0195"

1053. 2651

/gene="oppA_1"

/codon_start=1

/transl_table=11

/product="Oligopeptide Binding Protein"

/protein_id="AAD18348.1"

/db_xref="GI:4376466"

/translation="MRKISVGICITILLSVYLOGCKSSHSSTSRGELAINRDEP
BSLDPQVRLLEISLVKVIYEGVQENNSGNTEPALADYSLSDGLTYFKLSA
PNSGDLTAEDEIESKQVATQVSGIYAFALPNKVKIKQEGHLSIDHFGVHSPN
ESTLVLTSEPTSHFLKIALVFPVHKQSORTLOKSLPIASGAFPKNKOKOWIK
SKGHLHFDVAGTSMTIHFIPDNTAAKLNQKLNQKNGPWPGERIPOETLSNLQ
LPTNIHSEYPOKHOEAOQVAKKLEALQITAKDEHMLIPVSSASSLL
VOLIRHSEKESGAPVIGVEFALLQADLSNGNSLATGGWFADPMAFLIFAY
PSGVPPVAINKHDFTLEIQNEIQEQDHOKRSELVSQASLYLETFTHIEPIYHDAFQRA
MNKLSNLGVSPTGVVDFRYAKEN"

2802. 4385

/gene="oppA_2"

/note="Cp0196"

2802. 4385

/gene="oppA_2"

/codon_start=1

/transl_table=11

/product="Oligopeptide Binding Protein"

/protein_id="AAD18349.1"

/db_xref="GI:4376467"

/translation="MLRFAVFISTLWLTSGCSPQSSKGIFFVNNKMPRSLDFPK
TRLIADQTLRHLRYGLVEHSQNGEIPALAEYSITSEDGTYRTFKIKILNSGDP
LTAQDPSWKEILKEDASSVYLYAFILPKNRAIFDDTESPENLGVRLDKRLEIQ
LETPCAHLHFTLPLFPFVHETLRNYSFEMPTTCGAFRPSVLEKGLRLHLEKNP
MYNKSRLVKHLKIIIVQFI SNANTAAAILFKHKLDWQGPWGERIPPELSASLHQQDL
FSLPGASTTWLLENQKPKWNNKLSALSDAIDKMLTKVYQGLAEPTDHLHPLR
YPTGYPERKNERLLEAQOLFERALDEQMTREKLETTFTSTFSFGYRICOMLR
EOKVKYKFTIPVQGEFTIQQANLEGNSTVNWNTAAFIDPNSYLMFANPGGIS
PNYLQDSHTLLIKITOEKHKHLRNQILIEALDLEHCHILEPLCHPNRLNALNKNI
KNFLVRSITSDFRFIEKL"

4441. 5748

/gene="oppA_3"

/note="Cp0197"

4441. 5748

/gene="oppA_3"

/codon_start=1

/transl_table=11

/product="Oligopeptide Binding Protein"

/protein_id="AAD18350.1"

/db_xref="GI:4376468"

/translation="MFSRWITLFLFLISLTGCSYSKHKQSLIIPIHDDPVAFSPQ
AKRAMDISAQLLFDGLTRHRESNDLELAIASRYTVSEDFCSYTFEIKSALMSDG
TPITSEDIRNAMEYAQENSPHIOIFOGLNESTPSSNAIYHLDSPNDFPKLAFPAF
AIFPENPKLFSGPVTLVPEYPGNIIHKNNPNTYDHCVSNKSLIILIPDIYTAIH
LLNRKVDWVGQPHWGHOIBELHKOSQYHYTYTYPVEGAFLCLNTKSPHNLQNRHR
LATCIDRSIIIEALQGTQQAETLSRGAQPNQYKQKPLTPQEKLVLPVPSDILRC
ORIAEILKBQWKAAGIDILLEGLEYHLFVNKRKVQDYATATQTVGVVYVPGANLISEED
KILLQNEIPIVYLSVDYLTQDFIEGVYINASGAVDKLYTFP"

6032. 7612

/gene="oppA_4"

/note="Cp0198"

6032. 7612

/gene="oppA_4"

/codon_start=1

/transl_table=11

/product="Oligopeptide Binding Protein"

/protein_id="AAD18351.1"

/db_xref="GI:4376469"

/translation="MHLKPTLKSLLPMLLLTSSCKSKOKEPLGKHLVAMSHD
LADLPRLAYLSROASLAKALYEGLTRETDQGIALLAESAITYLSKDHVYTKLRPSV
WSDGPTLAYDFEKSILQLYEEFSPSIHTLLGVKNSSAIHNAOKSLETLGQIARDO
LTVLTLQEPFPFELTLIARVFSVPHRTRESYKKGTPPSTYISNGFVYKKEHQH
YLLLEKNPHYDVSVKLDVTKIIPIDASTATKLFKSKIDWIGSPWAPISNEDQK
VLSQEKILTVSVSTLLIYNLQKPLIQNKALRAIAHAIKSLILVPSGQEAUTL
VPPNLSQNLQKEISTEERTKARAYQKQKELTSEKLAELSILYPISSNSIIAQ
ETIROLKDTLGKIKIQGMEYHCFKLAKRROGDFATGCTGAETVSPVAFSLILGNR
DTOWRNSDYKTEKLYLPHAYKENLKRAMEIIEETPIIPILHGYKIYAIHPKION
TFGSLGHDTLKNIDLS"

7905. 8849

/gene="oppB_1"

/note="Cp0199"

7905. 8849

/gene="oppB_1"

/codon_start=1

/transl_table=11

/product="Oligopeptide Permease"

/protein_id="AAD18352.1"

/db_xref="GI:4376470"

/translation="MFSYIKNRITLNLISLWLTFLVYKTKIPGDPNDEGCNVLS
BEVLTLSRYGLDPLQYQYQYLSHAKTDFNSLYVDKRVNTIITAPPSAIL
GLOSIFISGGTALGIALAAKKKORRTILGASLIQISIPAFITLLQYVFAVKIP
LLPIACGSEFTHILPLALAVTPAFITQLTYSVSAALNKDYVLLAVAKGLSPKV
VIKHLPPVAFPTISYSAFTTITVTITQTAIENIFCIPGLGKWFICSIKQRDYPVALG
LSVFGYTLFMLSLLSDLSIQSIIDPQIRYAHGKKEKKR"

8880. 9734

/gene="oppC_1"

/note="Cp0200"

8880. 9734

/gene="oppC_1"

/codon_start=1

/transl_table=11

/product="Oligopeptide Permease"

/protein_id="AAD18353.1"

/db_xref="GI:4376471"

/translation="MENLSSPSSRWKSIIONKMLVGLTLLIILMALLGALLPWFYQ
DYETSGLKDLVSPCSRFPFGDTLGRCMFARTLGRLLSLIATLIATIDVQVGLIM
ATVAISGKKDILFMRTTEFLSLPRPIIILLVIFHGLLPLLIATLITWIPIS
RIIYQGLLLKPKFVLSAKAMHASTPHLLKHLHLLNTLPIIISILITFTINAIYEA
FIYPLGIGTQPPASLGITVKEGINAIDYIPWLPFFPSLIMIALISFNLIGEGAKTL
CLEGSHUG"

9727. 10581

/gene="oppD"

/note="Cp0201"

9727. 10581

/gene="oppD"

/codon_start=1

/transl_table=11

/product="Oligopeptide Transport AtPase"

/protein_id="AAD18354.1"

```

/db_xref="GI:4376472"
/translation="MDNYLNTKLTITSTNPKRTLENSLQKLNRLALVGEESG
KTTITKAILGFLPCLIKGTSLFEDITIKLSPKELHKIRGQKIATILQNAWGL
TPSMRIQMIIETIRGHRMKEEAYNKAMQLLDVCIIPNPKYSFSQYFELSGGMRQ
RVVTAIALSQPKLILADEPTTALDSMOAOLRILRNQOQKOATILLVTHNLSLVK
ELCNDICITKDGLIETGTVEIFLSPKHPYTLKLLNANVSKPIPKTSSPILKNKFPQ
LMSWOGSL"
10578..11366
/gene="oppP"
/feature="oppP"
/feature="CPH0202"
10578..11366
/gene="oppP"
/codon_start=1
/transl_table=11
/product="Oligopeptide Transport ATPase"
/protein_id="AAD18355.1"
/db_xref="GI:4376473"
/translation="MTLLSKDLSLIRGKILNHLNHLKGSVLIVGPSGSGKS
SLATLIDLLKPTTGITFHMDPKIPRAKQVQIWDIDSSLNPMCSIKGIISPLNI
IGTYSKAEONKETYNNVLNLPKYLHLPYKLSGQKRIARALVSPKELLICD
EPSSLDTLNQSLIDLFQTKKETQNTLFTIHDMSAAYIADTIAYMDOGSLVEHA
CREKIFSTPHHTTQDLDAIPFISLISSTEMPSEBYELOVASK"
BASE COUNT 3743 a 2799 c 1831 g 3275 t
ORIGIN

```

alignment_scores:

Quality: 2739.00 Length: 532
Ratio: 5.148 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-824-567-2 x AE001606 ..

Align seg 1/1 to: AE001606 from: 1 to: 11648

```

1 MetArgLysSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
|||||
1053 ATGGCGAAGATATCATGTTGGGAATCTGTATCACCATTCTCTTAGCCTCTC 1102

17 rValValGlnGlnGlyCysLysLysGluSerSerHisSerThrSerArgg 34
|||||
1103 CGTAGTCCTCCAAAGCTGCAAGGAGTCAGTCAGTCTCTCTATCATCTCGG 1152

34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
|||||
1153 GAGAACTCGCTATTATATAGAGATCAACCCGCTCTTTAGATCCAAAGA 1202

51 GlnValArgLeuSerGluIleSerLeuValLysHisIleTyrGluG1 67
|||||
1203 CAAGTCGCGACTCTTTTCAGAAATCAGCCTTTGTCAAACATATCTATGAGG 1252

67 yLeuValGlnGlnAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
|||||
1253 ATTAGTTCAGAAATAATCTTTTCAGAAATATAGAGCTGCTCTTCGAG 1302

84 luAspTyrSerLeuSerSerAspGlyLeuThrThrPheLysLeuLys 100
|||||
1303 AAGACTACTCTCTTTCTCGGACGACTCACTTATACATTTTAACTGAAA 1352

101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG1 117
|||||
1353 TCAGCTTTTGGAGTAAATGCGGACCCCTTAAACAGCTGAAGACTTTATAGA 1402

117 uSerTrpLysGlnValAlaThrGlnGlnValIleSerGlyIleTyrAlaPheA 134
|||||
1403 ATCTTGGAAACAAAGTAGCTACTCAAGAAGTCTCAGGAATCTATGCTTTG 1452

134 laLeuAsnProIleLysAsnValArgLysIleGlnGlnGlyHisLeuSer 150
|||||
1453 CCITGATCCCAATTAATAATGTCAGAAAGATCCAAAGAGGACCTCTCTC 1502

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValTh 167
|||||

```

```

|||||
1503 ATAGACCAATTTGGAGTGCATCTCTCTAATGAATCTACACTTGTGTAC 1552

167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184
|||||
1553 CTTGGAATCCCCCACTCGCATTTCTTAAAACTTTTAGCTCTTCCAGTCT 1602

184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
|||||
1603 TTTTCCCGTTCATAAATCTCAAGAACCCTGCAATCCAAATCTCTACCT 1652

201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpI1 217
|||||
1653 ATAGCAAGCGGAGCTTTCTATCTCTAAATAATCAAAACAAACAAATGGAT 1702

217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
|||||
1703 AAACTCTCAAAAACCTCTCCTACTATAATCAAAAGTCAGGTGGAAACTA 1752

234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
|||||
1753 AAACGATTACGATTCTCTTCTCCGATGCAAAACACACAGCAGCAAAACTA 1802

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgI1 267
|||||
1803 TTTAATCAGGAAAACTCAATTGGCAAGGACCTCTCTGGGAGAGCGCAT 1852

267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
|||||
1853 TCCTCAAGAAACCTCTATCCATTTTACAGCTTAAGGGCAGCTTACACTCT 1902

284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
|||||
1903 TTGATGTGCGAGAACCTCATGTGCTCACCTCAATATCAATAAATCCCC 1952

301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG1 317
|||||
1953 CTCACAAATATGAAGCTTAGAGAAGCCTTAGCATCAGCCTTAGATAAGA 2002

317 uAlaLeuValSerThrIlePheLeuGluValArgAlaLysThrAlaAspHisL 334
|||||
2003 AGCTCTGCTCTCACTATATCTTAGGCGGTGCAAAACCTGCCGATCATC 2052

334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
|||||
2053 TCCCTACCTCAAAATATCATAGCTATCCGAAACATCAAAACAAAGAGATG 2102

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG1 367
|||||
2103 GCACACGCCAAGCTTACGCTAAAAAACTCTTTAAAGAGCTTTAGAAGA 2152

367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
|||||
2153 ACTCCAAATCACTGCTTAAGAGCTCGAACATCTTAATCTTATCTTTCCCG 2202

384 alSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrp 400
|||||
2203 TTTCTCTGTCAGCAAGTCTTCTTACTAGTCCAACTTATACGAGAACAGTGG 2252

401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
|||||
2253 AAAGAAAGTTTAGGTTGCTATCCCTATTGTCGGAAGGAATTTGCTCT 2302

417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434
|||||
2303 TCTCCAGCAGACCTATCTTCAGGAACTCTCTCTTAGCTACAGGAGGAT 2352

434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
|||||
2353 GGTTGCGAGACTTTGCTGATCCTATGCAATTTCTAAGCATCTTTGCTTAT 2402

451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI1 467
|||||

```

2403 CCATCAGGAGTCTCCTCTATGCAATCAACCAATAGGACTTCTTAGAAAT 2452
 467 eLeuGlnAsnIleGluGlnGluAsnHisGlnIleGluLeuV 484
 2453 TCTACAACATAGAACAGAGCAAGATCACAAAACGCTCGGAATTAG 2502
 484 aIserGlnAlaSerLeuTyfLeuGluThrPheHisIleIleGluProIle 500
 2503 TGTCGACGCTTCCTTACTAGAGACCTTCATATATTATGACCGATC 2552
 501 TyrHisAspAlaPheGlnPheAlaMetAsnIleGluLeuSerAsnLeuG 517
 2553 TACCACGAGCATTTCAATTGCTATGATAAAAAAAGCTTCTAATCTAGG 2602
 517 yValSerProThrGlyValValAspPheArgTyrAlaIleGluAsn 532
 2603 AGTCTCACACAGAGATTTGAGCTTCCTGATGCTAAGGAAAT 2648

seq_name: gb_ba:AE002216

seq_documentation_block:
 LOCUS AE002216 11764 bp DNA linear BCT 30-MAY-2000
 DEFINITION Chlamydomophila pneumoniae AR39, section 44 of 94 of the complete genome.
 ACCESSION AE002216 AE0022161
 VERSION AE002216.2 GI:8163460
 SOURCE Chlamydomophila pneumoniae AR39.

ORGANISM Chlamydomophila pneumoniae AR39
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE 1 (bases 1 to 11764)
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 Nucleic Acids Res. 28 (6), 1397-1406 (2000)
 JOURNAL MEDLINE
 PUBMED 20150255
 REFERENCE 2 (bases 1 to 11764)
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

Direct Submission
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 On Jun 1, 2000 this sequence version replaced gi:7189484.
 Location/Qualifiers
 1..11764
 /organism="Chlamydomophila pneumoniae AR39"
 /strain="AR39"
 /db_xref="taxon:115711"
 /note="synonym: Chlamydia pneumoniae AR39"
 Complement(109..1707)
 /gene="CP0572"
 complement(109..1707)
 /gene="CP0572"
 /note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
 PID:40005; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=11

gene
 CDS
 product="peptide ABC transporter, periplasmic peptide-binding protein, putative"
 /protein_id="AAF38391.1"
 /db_xref="GI:7189485"
 /translation="MRKISVIGICTILLSVVLQCKSSHSSTSRGELAINIRDFP RSLDPROVRLISGLVHIEYGLVQNNLSGNIPEALAEYDLSLSDGLTYTFKLKSA FWSNGDPLAESDFISWQVATQVSGVIAFALNPKNVKRIQEGHLSIDFEGVHSPN ESTLVVTLESPTSHFLKLALPVFPVFKSQRTLQSKSLPIASGAFYFNKIKQKWIK

gene
 CDS
 complement(1653..2687)
 /gene="CP0573"
 /note="O-sialoglycoprotein endopeptidase; identified by match to PFAM protein family HMM PF00814"
 /codon_start=1
 /transl_table=11
 /product="O-sialoglycoprotein endopeptidase"
 /protein_id="AAF73688.1"
 /db_xref="GI:8163461"
 /translation="MLTLGLESCDETFACAIIVNEDKQILANIASODIHASGVVPE LASRAHLHIFQVINKALQANLLIEDMLIAVQTQGLIGSLVGVHFGKGIAGAK KSLIGVNHVEARHLYAAYMAQNVQFALGLVSGAHTAAFTIENPTSYKLIGTRDDA IGETFDKVRFLGLPYAGPLIEKLALGESDSYFPKAPVNYDFSGKLGKTRDYA IKGNSSPSPAPSEISLEKQRDIAASQKAACTTIAOKLPTIIKEFSRCLILGGGVA INEYFSATQTACNLVYFPFPAKLCSDNAIAGLGGENFQKNSSIDPEIRICAYQWE SVSPFSLASP"
 2771..3214
 /gene="CP0574"
 2771..3214
 /gene="CP0574"
 /note="similar to GB:M17532 SP:P15282 GB:X13968 PID:145356
 PID:43307; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="arginine repressor"
 /protein_id="AAF38392.1"
 /db_xref="GI:7189486"
 /translation="MKKVTIDEALKEILLEGAAQBELCAKLLAQGAFTQSSVSR WLKTKQAVKAGARGARYSLPSTKTTTRHLVLSIRHNASLIVIRTVPGSAWIAAL LQGLKDLBILGTLAGDDTFIVTPIDEGRLLPLMVSANLLQVFLD"
 3263..3916
 /gene="CP0575"
 3263..3916
 /gene="CP0575"
 /note="similar to SP:P10345 PID:41572 GB:U00096
 PID:1651370 PID:1787030; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="amino acid ABC transporter, permease protein"
 /protein_id="AAF38393.1"
 /db_xref="GI:7189487"
 /translation="MDHLAIARLLLRGCGYTCVSGIGILCGSILGLIGTIVTSLYF PSKLTLLANSYVTVIGCTPLFIQILIIYEGLEVLPIETPTPLVAGIILASMSAYL AENIRGGINSLSIGQESAMVGLYKQIFVFIYFQVFNILPLSUTNEFVSLIKSS ILMVGVPELTKVTKDIVSRELNPMMYLLICAGLYFLMTTTSFSCISRLSEKRRSYDN"
 3906..4586
 /gene="CP0576"
 3906..4586
 /gene="CP0576"
 /note="similar to SP:P41076 PID:624632 GB:U00096
 PID:1651271 PID:178570; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="amino acid ABC transporter, ATP-binding protein"
 /protein_id="AAF38394.1"
 /db_xref="GI:7189488"
 /translation="MTIRVRLNLSVNNKKILDGWTFSLRGHITLVFGKSGSKTMI LRALAGLVQPTQGDIIWEGEAPALVFOQPELFSHMTVLGKTHPQIHLKGRSTEAR KAFELHLLDIESEVAKNYPDOLSGGOKORVAIVESLCMDKHTLLEDEPTSLADPEFA SFRHLLETRDQELTVGLTITHTDMQFVHSCLDRIYLIDQGTVAGVYDKRDGELDSGHL SKYIHSQA"
 complement(4620..5978)
 /gene="CP0577"

CDS

```

complement(4620..5978)
/gene="CP0577"
/notes="hypothetical protein; identified by Glimmer2; putative"
/transl_table=1
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF38395.1"
/db_xref="GI:7189489"
/translations="MKRHSFDSSTKKAIVAKAIQIKIMETDPSLNVETNAEI
ESILQETKEIKQKLSQAEGLLEKYSQETLSNLENTNASIKLSIGSVTEELASLK
QLVESIEESIQGQDLOIQLVETISDKFLSSIGETLSNLDNQNVOGLIKENPE
KSEAASVYVOTLLEPKRGETHKVAVHDYVNSLQPHMSVAGRRFGHIDMNG
YRVLGEKNGEDAVSKDYLERVYSQLIDKVEDKPTKKNKLLYSOGTSPKLE
GFLGLLLISGSGFTWKSASNDGFPFSAHRKETESDTCQITSTLISGNQAG
PTKWSLKVLPVPSIFQIEKPEVQLSVLYSYEDWLPIDNIPNKSQPRIPALILGQTM
LAGQKDYLELAHOTNQTMLKSPNCSRFSLQLKOTNQFENSPVDYFIVHAHSCHS
Gf"
complement(6069..9497)
/gene="CP0578"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38396.1"
/db_xref="GI:7189490"
/translations="MKRSLWKILGICLSSIVIGLPIPLPOLLSTSGKLVFSLIH
KESGLSCSAEELKISWFGROTARKIKLTGKADKVEFSAEKPELDGSLILILVKKPKG
ITLGSNKLINEPASIDHPVSVDLPGSLITVYNCKILISEHGFITMKTVSGSLVS
GFYLEKSSKFTWKVSEDOQSNIFESVLSQAQFSSVPVAFKFIASFPF
WDHLLSYEDINLSAEATHNDKISMTAGEGNQIQKQCHIKSTFYIVGSSSF
IELKPELALCNQILIPSTPITSKQIHATVSVAKTIPDITWKHIEITSOAQLPEVA
IHPKDNALQIRKTLGKIKTEKFSIDRYSSSTVLGGASPSHLNLSIDNKKHLTK
FRUQAQLPHTVLRALFPOPEYVNPVDNAVYSLNTEGYKNNAHLEADALDNLPLKL
SCMSGAWKNFLFGOGTYHFVKNKQWELIPHSFYAEAFSCAKQITDNLFPKFSG
KITARENELLHAKGSPNEFIKPTETISILHGQCSPLSLVSNHLAPHLKLLTFS
FNDGSKFTYKNLQALLENPDYDLNLRILIPOLLISLDESSTSPSKDKIKQSGG
EIPSLSTITIKYQVRSPYFGSGDLNFVWYNPQDNKLLTSLNFKSALLGE
LKLVMDFSKSSSQGTLOWEVSERYAFKNASCPTCLLHRTANRLDLSKLC
PEETKGLCITLLAAGLGSLAETPLIFVDNVSKETRIINDFKGSRRANLDAKLEY
DLKQCLAPRODSKTLAEFLGEOVHLSPSPERFKQTAWIHIPSSFIAGIIPMSF
GLKAOISLAPRINVSINKAFREGGPDVIMVDSENLOAQIPLILNEKISILLRENLT
AHLINEDNKRALFOENPILLAGAYSQVPTLEIDKQNFYLPPIRPSFEFRQIAT
LDMKISTANTGTWYALFQELDTDQKQFESWFTPIFVSQKSLICKRXDALIDR
IRLAWGTDIAHDLFWTLGIDPEVIKKYFNTSLKTKNFKELIKIRGSISSPEVDWS
SAVARTALLKSVSLGNPFSSLDKLFSSLDGSTPPPTVHPFPWEKSNFDSIENK"
complement(9494..10786)
/gene="CP0579"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38397.1"
/db_xref="GI:7189491"
/translations="MFRKLPFSKKTGOKRLNRNGLLOAIQISKVLHLNEAKEA
CVLSYGLLTCVPLVPLFLRSLQHLNLNKEWLLIKFPDYKKPIVAIVERAYHATE
SNIGLVINGSFVFCWAGILMLLSLDGKNIKFRISWPTPSLKRVSFYVITLVSPMI
FIIVGSMYITQIMPLOYAKLSLSHSMYALFISRFVYLLIYALPCVAFIPRV
AIOKSWIITLQNGCFFIFLQVIVFQKAFSLQVSIYSFTYQALVAFSLFLLIYITMI
YLFQALFTIIONRGCTFIPLGKILPSCYLOLITSTVILATITROPNEGLSPIAQF
IAKQSKPIGVSVSOCLDLEKGFLEFPYNNGYQVPEFSELTINDIADKLLHREIFKK
FNPDLGITFIENSQNTFNQAKNENUTLSIARRIK"
complement(10849..11649)
/gene="CP0580"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38398.1"
/db_xref="GI:7189492"
/translations="MFRKLPFSKKTGOKRLNRNGLLOAIQISKVLHLNEAKEA
CVLSYGLLTCVPLVPLFLRSLQHLNLNKEWLLIKFPDYKKPIVAIVERAYHATE
SNIGLVINGSFVFCWAGILMLLSLDGKNIKFRISWPTPSLKRVSFYVITLVSPMI
FIIVGSMYITQIMPLOYAKLSLSHSMYALFISRFVYLLIYALPCVAFIPRV
AIOKSWIITLQNGCFFIFLQVIVFQKAFSLQVSIYSFTYQALVAFSLFLLIYITMI
YLFQALFTIIONRGCTFIPLGKILPSCYLOLITSTVILATITROPNEGLSPIAQF
IAKQSKPIGVSVSOCLDLEKGFLEFPYNNGYQVPEFSELTINDIADKLLHREIFKK
FNPDLGITFIENSQNTFNQAKNENUTLSIARRIK"
complement(10849..11649)
/gene="CP0580"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38398.1"
/db_xref="GI:7189492"
/translations="MFRKLPFSKKTGOKRLNRNGLLOAIQISKVLHLNEAKEA
CVLSYGLLTCVPLVPLFLRSLQHLNLNKEWLLIKFPDYKKPIVAIVERAYHATE
SNIGLVINGSFVFCWAGILMLLSLDGKNIKFRISWPTPSLKRVSFYVITLVSPMI
FIIVGSMYITQIMPLOYAKLSLSHSMYALFISRFVYLLIYALPCVAFIPRV
AIOKSWIITLQNGCFFIFLQVIVFQKAFSLQVSIYSFTYQALVAFSLFLLIYITMI
YLFQALFTIIONRGCTFIPLGKILPSCYLOLITSTVILATITROPNEGLSPIAQF
IAKQSKPIGVSVSOCLDLEKGFLEFPYNNGYQVPEFSELTINDIADKLLHREIFKK
FNPDLGITFIENSQNTFNQAKNENUTLSIARRIK"

```

gene

CDS

Glimmer2; putative*
/codon_start=1

alignment_scores:

Quality: 2739.00 Length: 532
Ratio: 5.148 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-824-567-2 x AE002216/rev ..

Align seg 1/1 to reverse of: AE002216 from: 1 to: 11764

1 MetArgLysLysSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
1707 ATGCGCAAGATATCAGTGGGAATCTGTATCACCATTCTCTTACCTCTC 1658

17 rValValLeuGlnGlyCysLysGluSerSerHisSerSerThrSerArg 34
1657 CGTAGTCTCCAGGCTGCAAGGAGTCCAGTCACTCCTTACATCTCGGG 1608

34 LyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
1607 GGAACCTCGCTATTATATAGAGATGAACCCCTCTTTAGATCCAGA 1558

51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGlu 67
1557 CAAGTGGGACTCTTTTCAGNAATAGCCCTGTCAACATATCATATGAGG 1508

67 yLeuValGlnGlnAsnLeuSerGlyAsnIleGluProAlaLeuAla 84
1507 ATTAGTTCAAGAAAATAATCTTTAGGAAATATAGACCTGCTTTTCAG 1458

84 LuAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeu 100
1457 AAGACTACTCTCTTCTCGGACGACTCACTTATCTTTTAAACTGAAA 1408

101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117
1407 TCAGCTTTTGGAGTAAATGGCACCCCTTACAGCTGAAGACATTATAGA 1358

117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPhe 134
1357 ATCTTGAAACAAAGTAGCTACTCAAGAAGTCTCAGGAATCTATGCTTTG 1308

134 LaLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
1307 CCTGAATCAATATAAAATGACGAAGATCCAAAGAGGAGACACCTCTCC 1258

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValVal 167
1257 ATAGACCACTTTGGAGTGCCTCTCTTAATGAATCTACACTTGTGTGTTAC 1208

167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProVal 184
1207 CTTGGAATCCCAACCTCGCATTTCTTAAAACTTTTAGCTCTTCCAGTCT 1158

184 HePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
1157 TTTTCCCTCTCATAAATCTCAAGAACCCCTGCAATCCAAATCTCTACCT 1108

201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrp 217
1107 ATAGCAAGCGGAGCTTCTATCTTCAAAATATCAAAACAAACAAATGAT 1058

217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThr 234
1057 AAAACTCTCAAAAAACCTCTACTACTATATCAAGTCAAGTGGGAACATA 1008

234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
1007 AAACGATTACGATTCACTTCACTTCCCGATGCAACACACAGCAGCAAACTA 958

```

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArg11 267
|||||
957 TTTAATCAGGGAACCAATCAATTGGCAAGACCTCCTTGGGGAACGCAAT 908
|||||
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
|||||
907 TCTCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGCACTTACACTCTT 858
|||||
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
|||||
857 TTGATGTCGCGAGCACTCATGGCTCAGCTTCAATATCAATAAATTCCTCC 808
|||||
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG1 317
|||||
807 CTCACCAATATGAAGCTTAGAAGCCTTAGCATCAGCCTTAGATAAGA 758
|||||
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
|||||
757 AGCTCTTGCTCAACTATATCTTAGCGCGTGCAAAACCTGGCGATCATC 708
|||||
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
|||||
707 TCTTACCTACATAAATATTCATAGCTATCCGGAACATCAAAACAGAGATG 658
|||||
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG1 367
|||||
657 GCACACGCCCAAGCTTACGCTTAAAAAAGCTTTTAAAGAGCTTTAGAAGA 608
|||||
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
|||||
607 ACTCCAAATCAGCGCTAAGATCTCGACATCTTAATCTTCTTCTTCCGG 558
|||||
384 alSerSerAlaSerSerLeuValGlnLeuLeuArgGluGlnTrp 400
|||||
557 TTTCTCTCGTCAGCAAGTCTTTTACTAGTCCAACTTATACGAGACAGTGG 508
|||||
401 LysGlnSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
|||||
507 AAACAAGATTAGGTTCGCTATCCCTATGTCGGAAGGAATTTGCTCT 458
|||||
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434
|||||
457 TCTCAAGCAGACCTATCTTCAGGGAACCTTCTCTTAGCTACAGAGGAT 408
|||||
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
|||||
407 GGITCGCAGACTTCTGCTATGCTATGCAATTTCTAAGCATCTTTGCTTAT 358
|||||
451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI1 467
|||||
357 CCATCAGGAGTTCCTCCTTATGCAATCAACCATAGGACCTTCTTAGAAT 308
|||||
467 eLeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuV 484
|||||
307 TCTACAAAACATAGAACAGAGCAAGATCACCAAAACGCTCGGAATTAG 258
|||||
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
|||||
257 TGTCCGACGCTCTCTTTACCTAGAGACCTTTCAATATTATTAGCCGATC 208
|||||
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG1 517
|||||
207 TACCACGCGCATTTCAATTTGCTATGAATAAAACCTTTCTAATCTAGG 158
|||||
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
|||||
157 AGTCTCACCAACAGAGTTGTGGACTTCGCTTATCTCTAAGGAAAT 112

```

seq_name: gb_ba:AP002545

seq_documentation_block:

LOCUS AP002545

DEFINITION Chlamydomydia pneumoniae J138 genomic DNA, linear BCT 08-DEC-2000

complete sequence,

section 1/4.

AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
 AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
 AB036074 AB036075 AB036076 AB036077 AB036078 AB036345 AB036346
 AB036347 BA000008
 AP002545.2 GI:9956082

Chlamydomydia pneumoniae J138 (strain:J138) DNA.

Chlamydomydia pneumoniae J138

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.

1 (sites)

Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
 Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
 Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA

Nucleic Acids Res. 28 (12), 2311-2314 (2000)

20330349

2 (bases 1 to 299650)

Shirai,M.

Direct Submission

Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
 School of Medicine, Department of Microbiology; 1-1-1
 Minamikogushi, Ube, Yamaguchi 755-8505, Japan

(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
 Fax:81-836-22-2415)

On Aug 31, 2000 this sequence version replaced gi:6172286

gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318

gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162

gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422

gi:8547426 gi:8547430 gi:8978373

AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)

AB033845-AB038347: Submitted (14-Feb-2000)

AB036071-AB036078: Submitted (18-Dec-2000).

FEATURES
 source

1..299650

/organism="Chlamydomydia pneumoniae J138"

/strain="J138"

/db_xref="taxon:138677"

/note="synonym:Chlamydia pneumoniae (strain J138)"

complement(1..282)

/gene="CP10001"

complement(1..282)

/gene="CP10001"

/codon_start=1

/transl_table=11

/product="CT001 hypothetical protein"

/protein_id="BAA98211.1"

/db_xref="GI:8978374"

/translation="MLGKTIIRGLSLIVLCALNVGLIGITHNKLNIIAKLCGGVSTP

ATQITYIIIGIAGVICLLSFCPCFKSKSRHSHGSDSCSGGCHSHSDKN"

573..878

/gene="gatC"

573..878

/gene="gatC"

/codon_start=1

/transl_table=11

/product="Glu-tRNA Gln amidotransferase (C subunit)"

/protein_id="BAA98212.1"

/db_xref="GI:8978375"

/translation="MEQFHLDRREITLLAKASALQSEELIOEYOTLSAVITSMKEA

LAIEDDADSCESLFMHVNVNVEDLREDSVTSDFNREEFLRNVPESLGLYKVPATK"

895..2373

/gene="gata"

895..2373

/gene="gata"

/codon_start=1

/transl_table=11

/product="Glu-tRNA Gln amidotransferase (A subunit)"

/protein_id="BAA98213.1"

/db_xref="GI:8978376"

/translation="MYRYSALELAKAVTILGELTATGVTQHFFHRIIEAAGVCAFTSL

CKQALAEQLIDKKRSRGEPLGKAGVPVGIKDNHTVGLTKTCASRVLENQPPFD

ATVVERIKKEDIILGLKLNDEFANGSTTILYSNFHTHPNWDLSRVPGSGSGGSAAY

233905 GAGAACTCGCTATTAATAAGAGATGAACCCGCTTTTAGATCAAGA 233954
 51 GlnValArgLeuLeuSerGluIleSerLeuValHisIleIleYrGluI 67
 233955 CAAGTGGAGACTCTTTCAAGAAATAGCCTTGTCAACATATCTATGAGGG 234004
 67 yIeuValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
 234005 ATTAGTTCAAGAAATAATCTTTCAGGAATATAGAGCGTCTCTGAG 234054
 84 LuAspTyrSerLeuSerAspGlyLeuThrYrThrPheLysLeuLys 100
 234055 AAGACTACTCTCTTCTCGGAGGACTCACTATATCTTTTAAACTGAAA 234104
 101 SerAlaPheThrPsrAsnGlyAspProLeuThrAlaGluAspPheIleG 117
 234105 TCAGCTTTTGGAGTAAATGGGACCCCTTAACAGCTGAAGACTTTATAGA 234154
 117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
 234155 ATCTTGGAAACAAGTAGCTACTCAAGAACTCAGGAATCTATGCTTTTG 234204
 134 laLeuAsnProIleLysAsnValArgLysIleGlnGluValSerGlyIle 150
 234205 CCTTGAATCCATTAATAATGTCAAGAAATCCAGAGGGACACCTCTCC 234254
 151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
 234255 ATAGACCATTTTGGAGTGCATCTCTTAATGAATCTACACTTGTGTTAC 234304
 167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184
 234305 CCTGGAAATCCCAACCTCGCATTTCTTAAAACTTTTAGCTCTCCAGTCT 234354
 184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
 234355 TTTTCCCTTCATTAATCTCAAGAACCCCTGCAATCCAAATCTCTACCT 234404
 201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpI 217
 234405 ATAGCAAGCGGAGCTTCTATCTTAAATAATATCAAAACAAAACAAATGAT 234454
 217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
 234455 AAACCTCTCAAAAACCCCTCACTACTATATCAAAAGTCAGTGGAAACTA 234504
 234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
 234505 AAACGATTACGATTACCTTCATTCCTGATGCAACACAGCAGCAAAACTA 234554
 251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgII 267
 234555 TTTAATCAGGGNAACACTCAATGGCAGGACCTCTCTGGGAGAAAGCAT 234604
 267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
 234605 TCCTCAAGAAACCCCTATCCAAATTACAGTCTAAGGGCAGCTTACACTCTT 234654
 284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
 234655 TTGATGTCGAGGAACCTCATGCTCACTTCAATATCAATAAATCCCC 234704
 301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGI 317
 234705 CTCACAATATGAAGCTTAGAAGAGCTTAGCATCAGCTTAGATAAGGA 234754
 317 uAlaLeuValSerThrIlePheLeuGlyArgAlaIleYrThrAlaAspHisL 334
 234755 AGCTCTGTCTCACTATATCTTAGGCGTGCAGAAACCTGCCGATCATC 234804
 334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
 234805 TCCTACCTACAAATATTCATAGTATCCCGACATCAAAACAAAGAGATG 234854

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluI 367
 234855 GCACAAACGCCAGCTTACGCTAAAAAATCTTTTAAAGAGCTTTAGAAGA 234904
 367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
 234905 ACTCCAAATACCCCTAAAGATCTCGAACATCTTAATCTTATCTTTCCCG 234954
 384 alSerSerAlaSerSerLeuValGlnLeuIleArgGluGlnTrp 400
 234955 TTTCTCTCGTACGCAAGTCTTTACTAGTCCAACTTATACGAGAACAGCTGG 235004
 401 LysGluSerLeuGlyPheAlaIleProlleValGlyLysGluPheAlaLe 417
 235005 AAAGAAATTTTAGGGTTGCTATCCCTATTGTCGAAAGGAATTTGCTCT 235054
 417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyLyt 434
 235055 TCTCCAAACGACAGCTATCTTCAGGGAACTTCTTTAGCTACAGGAGGAT 235104
 434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
 235105 GGTTCGAGAGCTTGTGATCTCTATGGCATTTCTAACGATCTTTGCTTAT 235154
 451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI 467
 235155 CCACTAGAGTTCCTCTTATGCAATCAACCATAGGACTTCTTAGAAAT 235204
 467 eLeuGlnAsnIleGlnGlnGlnAspHisGlnLysArgSerGluLeuV 484
 235205 TCTCAAAACATAGACAGACAGATCACCANAACGCTCGGAATTAG 235254
 484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
 235255 TGTGCAAGCTTCTCTTTACCTAGAGACCTTTCATATTATTAGCGCATC 235304
 501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuGI 517
 235305 TACCACACCCATTTCAATTTGCTATGAATAAAAAAATTTCTTAATCTAGG 235354
 517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
 235355 AGTCTACCAACAGAGATTTGAGCTTCCGTTAAGTAAGGAATAAT 235400
 seq_name: gb_ba:AE001293
 seq_documentation_block:
 LOCUS AE001293 11944 bp DNA linear BCT 30-OCT-2000
 DEFINITION Chlamydia trachomatis section 20 of 87 of the complete genome.
 ACCESSION AE001293 AE001273
 VERSION AE001293.1 GI:3328597
 KEYWORDS
 SOURCE Chlamydia trachomatis.
 ORGANISM Chlamydia trachomatis
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 11944)
 AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
 Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
 Koonin,E.V. and Davis,R.W.
 Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis
 Science 282 (5389), 754-759 (1998)
 JOURNAL 99000809
 MEDLINE 9784136
 PUBMED
 REFERENCE 2 (bases 1 to 11944)
 AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
 Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
 Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 Nat. Genet. 21 (4), 385-389 (1999)
 JOURNAL 99206606
 MEDLINE 10192388
 PUBMED
 REFERENCE 3 (bases 1 to 11944)

AUTHORS

Stephens, R. S., Kalman, S., Lammel, C. J., Fan, J., Marathe, R.,
 Aravind, L., Mitchell, W. P., Olinger, L., Tatusov, R. L., Zhao, Q.,
 Rooin, E. V. and Davis, R. W.
 Direct Submission
 Submitted (20-MAY-1998) Program in Infectious Diseases, University
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA

FEATURES

Location/Qualifiers
 1..11944

/organism="Chlamydia trachomatis"

/strain="D/UW-3/CX"

/db_xref="taxon:813"

complement(105..878)

/gene="CT192"

complement(105..878)

/gene="CT192"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAC67784.1"

/db_xref="GI:3328598"

/translation="MFFSGVETLPSTCLISKKTRKEEFVRSVQGEASQNTLSWTIR

PRPSIQGKESVLAIEVIGTVAIAAGACAAAGGAFSGVIGLVFLGVVLAATGLIL

AVLEFCHIRSMSEKVIYLTQKDLFKEVIEQATPLIEFASYTCPCGIPLSGPEEQV

QERVLQKDLSDSHAPKYIAVGSHVVELVAKGIGRNGERLLEBGI DTGDNFVRWD

ELFLTGQGMVRLDFCCVKLPKKSSESINNVSDC"

complement(905..2023)

/gene="tgt"

/note="CT193"

complement(905..2023)

/gene="tgt"

/codon_start=1

/transl_table=1

/product="Queuine tRNA Ribosyl Transferase"

/protein_id="AAC67785.1"

/db_xref="GI:3328599"

/translation="MALRFEILHQSKSRARVGRIGRTAHGYIDTPAFVPVATGALKG

VLDSNIPMLFNTYHLIVHPGAEATAAGGLHOFI GRNAPIITDSGGFOLFSIAYGS

VADEIKSCGKKGGNTIIKVNDGVEHFKSVROGRKLFLSPSLVQAQKDLGADIIPL

DELLPDAFTVHOSORTVYWEKSLDYHLKNGIQSMYGVIHGGTFPDQRKLGCK

FVELDFPDGSAIGGSLGNLQDIVEVVGVTANLSAERPHLLGIGDLPISWATVGF

IDSDSYTPAARHGMLITSQGLKINNVORYSDNPIEPFGCSCLACSGGITRAYLR

HLFKVHEPNAGIWAISHNMHHQKVMKREIREGLNDRI"

complement(2150..3562)

/gene="tgtE"

/note="CT194"

complement(2150..3562)

/gene="tgtE"

/codon_start=1

/transl_table=1

/product="Mg++ transporter (CBS Domain)"

/protein_id="AAC67786.1"

/db_xref="GI:3328600"

/translation="MDSKTSHLDELCKLEBAFTTLFAGHESODLTSIVSVNPIDL

AVASCLSDRSILYKNLOSIAKFAFIINTDSASRWAFIRNLSDEICALTEQMP

DEAIWLDDIPDRYRILDLIDKALKIRDLQKHGNTAGRLMTEFFAFMETTV

KEVATCIENFGIDLRLVFLVDFKGLQFVDSRLIIASPEMLKQIMRPIEHLV

ADTTEEVVDLVERKYAVLPVVDENELIGALTYEDVETIEDIADETIARMAGTTE

DVYHDCWVORTILRAPWLLITLCLAGLVSVMAYQKIAPIILLAWIFFIPLVNG

SGNVGVCSTILVRSMATGILSGRRRETILKEMSIGLLTGVALGLOGLVWCMGCL

GLGLPATGVOLGVTVSVGILGASLTATLGLVSPFFFAKTGVDPALASPIVTLND

INSMWIFLLITGLNLVLFK"

complement(4001..5092)

/gene="CT195"

complement(4001..5092)

/gene="CT195"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAC67787.1"

/db_xref="GI:3328601"

/translation="MWSMLNLPPEAEVLRPVTAKSPNNSAWRATQKRTGCSLNAS

STRPSSARVTSFVGVVLVGLVLCFRIAYLANSRVLTFPRKFAIILPILFGGLAI

LFRTAGKVDVLYGKIQPFIISRQWVERVILCEKEGEFIRPIQEPDMYMDVSLDRTGSG

gene

CDS

IAPVVTPMDARTVICWTSILLPIFISIVRMLYNIPREFFIVPFYILFQWVRNXYOTD
 IKPERFVCSDIVMERISLQAVKAPFYGAVCYLANLYGLINPLISCRVNLASLERDW
 NNDVRSRGWVIECEKNYFEGSGTRSGLGQNAWYLLGCPQVQLFLLADGVIISGA
 RSIQSPPSEKEYLASFLYGAVPRLAGF"

5316..5636

/gene="CT196"

5316..5636

/gene="CT196"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAC67788.1"

/db_xref="GI:3328602"

/translation="MATVTPNPNYSLPFCEKVMVSSATLAASHFFNHKILIPILVGYFC

ILLGALILTGVTITPISIAASVFLSLGVLIVAGLCAAFKAPLFSVTSKASTLLH

TSCD"

5712..6728

/gene="gcp_1"

/note="CT197"

5712..6728

/gene="gcp_1"

/codon_start=1

/transl_table=1

/product="o-Sialoglycoprotein Endopeptidase"

/protein_id="AAC67789.1"

/db_xref="GI:3328603"

/translation="MLTLGLESCDSCSLVQNGKTLANKIASODIHASYGCVIPEL

ASRAHLQTFPELLTAATQSAGVSLIEDIELISVANTPGLIGALSIGNFRANGSLAKR

PLIGNVHVEAHLVYACMEAPATQFFALGLAISGAHTSLFLMPDATTFLLIGKTRDAIL

GETDKVARELGLPYFGGKLEELADGADAFAPSPARYSGYDFSGLKATVIAL

KGNSSAKAPPEVSETKRNIAASFQKAVMTIAQKLPDIVKTFSCESLI VGGGVAN

NSYFRLLNQICSLPIYFPFSQLCSDNAAMIAGLGRLCNTRHVSKEVIPCARVQWE

SACS"

6692..8248

/gene="oppA_3"

/note="CT198"

6692..8248

/gene="oppA_3"

/codon_start=1

/transl_table=1

/product="Oligopeptide Binding Protein"

/protein_id="AAC67790.1"

/db_xref="GI:3328604"

/translation="MRKISVGICLLILALATSGCKSSSNATHRSRPATHVAVSVKDDP

WTFDPREVRLISDNLIIHLVYGLVQETPSGEVPEPALAESFSLSEDKTYTNLKKAF

RTSGDLITADHVRVSNWDLONRVASLYSFAFLPIDVWKSDFFAKDDHTLVINLLTP

TPFLKLLTLPPFYVPHSQHQRKEKSLPSTGAFFLAEKKDRWLKLEKSPYYNK

DQVAVQICIIIPDQQTASALFNOGKLDWGLPWGHSIPQETLATTNKRPRSFDI

SGTSMVTFNTAKKPFSSKRLQALSLVNLKEALASLAFYKPAKHLPAHLHTYPEQS

YKQOEAITLAKSLLEALTILNMTIEDLEKYLPTFSATSMNSQIAQMLRQWRISLG

ITPFCGKEYALLQNDLIGNTFMSIGWFAFSDPLAFLSFSSKGVKPYALQDPQF

DQLILSTETKPNQKRSALISEASLYIERQNVIEPLYVDFVHYTNNKLSFVRLHPSG

LVDNRVAKNS"

8407..9348

/gene="oppB_1"

/note="CT199"

8407..9348

/gene="oppB_1"

/codon_start=1

/transl_table=1

/product="Oligopeptide Permease"

/protein_id="AAC67791.1"

/db_xref="GI:3328605"

/translation="MLSVIKRRLFLNLSLWVVTLLTFELIKTPGDPFNDENGNIIS

SETLALKNRYGLDKPLFTQVLYLKLCLTLDFGSLYKORTVISAALPSSAIL

GLSICLSLFGGITGLILAFYKSCGRTIFFSSVQISVFAVIGAFIYVFAIKYS

CLPIACWGNFSTLPSLTAIIPMAFITQLTCAVSANLKKDYVLAYAKGLSPFKV

LKHLPLAFVPIYSYSAFITMTITMTFTSIENLFCIPGLCKWFCISKQORDPYPIIG

LSVFGAFMLTSLCCDLQAQWIDPQIRVSYGKRSK"

9360..10225

/gene="oppC_1"

/note="CT200"

9360..10225

gene

CDS

gene	<code>/db_xref="GI:7190507"</code> <code>/translation="MDSSCALKNLSSCEVSYPEVIATRKREVALVILVIGTFLAAIGGACAVATSGWSTGIGTAVGLATGFGFLALMEICRAVYGRQOKIQIQEQLVSSPIREATLPIEQAVSHPELPQLSQSEVRCRHPRIILORGSKPAETPKFIATVGNVSVELVAVGMIGPMGRGNVPGQTLVRLDELFGALGRMSELVRLDGFCCVKLVATLGDITSVSKTSADRPASSE"</code> <code>complement(1227..2345)</code> <code>/gene="TC0465"</code> <code>complement(1227..2345)</code> <code>/gene="TC0465"</code> <code>/note="similar to GB:AL009126; identified by sequence similarity; putative"</code> <code>/codon_start=1</code> <code>/transl_table=11</code> <code>/product="queuine tRNA-ribosyltransferase"</code> <code>/protein_id="AAF39315.1"</code> <code>/db_xref="GI:7190508"</code>
CDS	<code>/translation="NALPEFILLHQSCKSRARYGRIETHEGICIDTFAFVPAVATNALKGVLDHNSIPLMFCNTYHLIVHFGAEVAVMGGUHQFIGNAPIITDSGGFQIFSLAYGSAEIECKGKKGKGENSIKIINDEGSAWFKSLYDGRKRLFSPEVSQAQKHLGADIITPLDELLPFDPAFYQSSORTYAWEKSLYDHLANPGYSQMYHGHTGTFPQWRKLCQFVEDLFPDASATGSGSLGNLRDITGVVDVYTTANLSIERPHLLGIGDLPLSATVGVGQIFSDSSYPTKAARGHMLTSSQSLKINNOYASDLNPIEPGCACPACSQGISPAVLRHLFVKEHPNAGIWSIHNMYMOKIMSKIREKLNDQL"</code> <code>complement(2467..3939)</code> <code>/gene="TC0466"</code> <code>complement(2467..3939)</code> <code>/gene="TC0466"</code> <code>/note="magnesium transporter; identified by match to TIGR protein family HMM IGR00400"</code> <code>/codon_start=1</code> <code>/transl_table=11</code> <code>/product="magnesium transporter"</code> <code>/protein_id="AAF73559.1"</code> <code>/db_xref="GI:8163227"</code>
gene	<code>/translation="MGQNEKEKLDVALTSGNLMSKSHLDDSEKLEAEAPDLITAGEHSDLLSTIVSAVNPIDILAYAVSCLPSDSRSILYKSLDSTASKIAFLINDASRWALFRNLSGEICALLIQMPDPEAIWVLDIDIPORRYRILDLIDVKKALKINDLQKHGRNAGRLMNEFFALMETTYKVEATCIRNNGVDLTRNFVDLDFKGEIQGFVTDRLSIASPEPLQKIMRPMVEHKVLADITREEVVDLVERVYKAVLPVVDENELGAIYTDVDEVETIDIADETIAAGVMDGTEVGYDCHVQVRFLLRPALLITLCAGILVASVMAVYQKIAPTLIAVVFIFPLVNGSLGVQOCSITLVRSMATGTLSFGRRETILKEMSLGILTPVAGLILGLVACVCCGCLGLGSTGGVGLGVYVSGILGASUTATILGLVLPSPFFAKIGVDPALASGPITVALINDIMSIFLLIITGTLNLVFFK"</code> <code>4208..4366</code> <code>/gene="TC0467"</code> <code>4208..4366</code> <code>/gene="TC0467"</code> <code>/note="hypothetical protein; identified by Glimmer2; putative"</code> <code>/codon_start=1</code> <code>/transl_table=11</code> <code>/product="hypothetical protein"</code> <code>/protein_id="AAF39316.1"</code> <code>/db_xref="GI:7190509"</code>
gene	<code>/translation="MDKAHTVHFHPQLQYDEFTVDLQAKSLGLLKTVAAYAKLAASQPSQSGSPI"</code> <code>complement(4328..5410)</code> <code>/gene="TC0468"</code> <code>complement(4328..5410)</code> <code>/gene="TC0468"</code> <code>/note="conserved hypothetical protein; identified by Glimmer2; putative"</code> <code>/codon_start=1</code> <code>/transl_table=11</code> <code>/product="conserved hypothetical protein"</code> <code>/protein_id="AAF39317.1"</code> <code>/db_xref="GI:7190510"</code>
gene	<code>/translation="MFTNLPSAEVRLRPVTTSEFPASSKATOKRITISCTLTNPFYKTSLTRIVCFVGVILGICRIYALANAGLLFSPKIALALSILVLLGGGIVYCKIARKIDLVYGEKTOPFAIRKREQVILCEKQOIRIIPQDPMDSLDLQKQSGIAPVETTPAKARTACITACILIPITVVRVLYNAFFRLPIYVFMQILQYEDLPBEOYQCTDIFREMSRFSQVAKPPYVACVLYLASLYGLINPLSGRIVMASVMDWND</code>
CDS	

```

VIRSGINGIFCEKNCLLEGGTRSGLGQAHAWLLGCFQFRLFLKKGGEIYSGARPS
5581. 5946 /gene="TC0473"
IQAFPEKNEYLTSLYGAALGRITSR"
5581. 5946 /gene="TC0469"
5581. 5946 /gene="TC0469"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39318.1"
/db_xref="GI:7190511"
/translation="MNLFSLNNQKARKMOLTIVNTNSRIFPFSNERILPGSTILSTD
SFVSFKILPLLAGSICLLGALLAGTITVTPAIAASFLSLGVILVAGIGLCSV
RPPLHLSINQSDLHIYY"
6012. 7034 /gene="TC0470"
6012. 7034 /gene="TC0470"
6012. 7034 /gene="TC0470"
/note="O-sialoglycoprotein endopeptidase; identified by
match to PFAM protein family HM PF00814"
/codon_start=1
/transl_table=11
/product="O-sialoglycoprotein endopeptidase"
/protein_id="AAF73560.1"
/db_xref="GI:8163228"
/translation="MLTGLSSQDETSCALVENKGLIANRIASQDIHAAYGGVPIPEL
ASRAHLQIPKLLAAVAQDAEVSLEDVLLSVANTPGLIGALSGVNFPAKGLASGLKK
TLIGNVHVAHLVYACLESPIRPPALGALSGAHTSLFMPNATVFLILGKDDAI
GETDKVAREFLGLPFGQKLEELAQDDEAFYFRAKVSQNDFFSGLKATVLYAL
KGNSSAKAPFVSETQRNIAASFQKAAPMTIAQKLDIVKAFSCSLVVGQGVAN
NRYFRLNNQTSCLPTYPFSSQLASQNAAMIAGLERLFCNQTVYSKEVIPCARYQNE
SACLSH"
6992. 8554 /gene="TC0471"
6992. 8554 /gene="TC0471"
6992. 8554 /gene="TC0471"
/note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
PID:40005; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein, putative"
/protein_id="AAF39319.1"
/db_xref="GI:7190512"
/translation="NRKVSVCILLLIALATAITGCSKSSSNKSHSSNQSQSVSMKD
DPTDFPREVRLSDINLIHLYEGVQETPSGEVPEPALAESFELSEKTYTFHLKK
ALMNSGDLITAHDFVRKNDVLQNRASIVSFAPIPLDLSQCGFFAKDNHTLVNLH
TPTPHFLKLTLPVYPVPHPEQIRNEAKALPISTGAFCLKEKDRRWIKLEKNYY
NKQVAIQEIHVIPDOQTASALFKQGLDMQGLPWGHSIPQALATANKRRTPQSF
DISGTSWLFNTSKVPFSPKRLQALSLVINKELASPTFVKPAKHLPLHLTYPEQ
PTYQOEAVILAKTLQBLATDLNMTIKLEKPLIFSATSVNSQMAQMRDQWRRI
LGTFPICGEYALQNDLTNTFMSINQWFAFSDPLAFLSVFSKGLKPYALQDP
LFQDLILSTETBKDPKRISLISEASLYLEKQNIIEPLYHDFVHYAANKLSFVRLHP
SGLVDMRYAKNS"
8668. 9612 /gene="TC0472"
8668. 9612 /gene="TC0472"
8668. 9612 /gene="TC0472"
/note="similar to SP:P24138 GB:X56347 PID:580897
GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF39320.1"
/db_xref="GI:7190513"
/translation="WLGIKKRLLFNLLSLWVVLVTLFTFIITIPGDPFNDENGLVS
EETLALNKRVRGLKPLFTQYLIYLCCLTDFGESLIYKDRVTIGITLAFPSAIL
GLSICLALFGGTLHAAAFYKGGRTIFFSIIQISVPFAVIGAFQIYFAIKYS
LPIACWDSFHTLPSIALAIAPATITQLYTASVSKLKDQVLLAYAKGLSPKLV
LKHVLYALPEVYSYAFLLITLTMTGTFSTENLCTIPGLGKFWFICIKQORDYPTMLG
LSVYGAFFMLISLCLDLQAWIDPQIRYSYKSHKQ"

```

```

gene 9618. 10481 /gene="TC0473"
CDS 9618. 10481 /gene="TC0473"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF39321.1"

alignment_scores:
Quality: 1345.00 Length: 538
Ratio: 3.135 Gaps: 9
Percent Similarity: 79.740 Percent Identity: 49.442

alignment_block:
US-09-824-567-2 x AE002315 ..

Align seg 1/1 to: AE002315 from: 1 to: 12173

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6992 AUGCGCAAGGTATCAGTGGGAATCTGC.....TTGCTTATCGGRTAGC 7035

17 rValValLeuGlnGlyCysLysGluSerSer.....His 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7036 AACAGCAATACCGGCTGCTCCAAATCTCTCAACAAATCAAACTACT 7085

29 erSerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArg 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7086 CCTCCTCTPAATCAA...AGCGTATCTGTGAGCATGAAGGACGATCCTCG 7132

46 SerLeuAspProArgGlnValArgLeuSerGluIleSerLeuVally 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7133 AGCTTTGACCTCGAGAGGTAGGCTTCTTCGGATATTAAATTGATCCA 7182

62 shisIleTyfGluGlyLeuValGlnGluAsnLeuSerGlyAsnIleG 79
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7183 CCATCTTTATGAGGACTGGTGGTCAAGAGACTCCT...TCGGGAGAGTTT 7229

79 luProAlaLeuAlaGluAspTyfSerLeuSerSerAspGlyLeuThrTy 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7230 TCCCTGCATTAGCAGAAAGTTTCTCTTATCTGAAGATAAGAAACCTTAC 7279

96 ThrPheLysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAl 112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7280 ACTTTCACCCTAAAAAAGCACTCTGGAGTAATGGAGACCTGATTACTGC 7329

112 aGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSerG 129
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7330 TCATGATTTGTACGTTCTTGAAGCATGTTTTCGCAAAATCCGATTGCAA 7379

129 lyIleTyfAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGln 145
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7380 GTATCTACTCTTTCGCTTTTACCATT..... 7408

146 GluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSe 162
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7409 .....GATTTAAGTAAAGAT...TGTCGATTTCTCGCTAAGGACAAATCA 7449

162 rThrLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeuL 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7450 TACACTGTGTTACAAITTCATACCCCTACTCCACACTTTCTGAAATATAC 7499

179 euAlaLeuProValPhePheProValHis...LysSerGlnArgThrLeu 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7500 TTACCTCTCCAGTATTTTATCTCTGACCCCGCAACATCAATTCGGAAT 7549

195 GlnSerLysSerLeuProIleAlaSerGlyAlaPheTyfProLysAsnIl 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7550 GAAGCAAAAGCTCTTCCCAATATCTACAGGGGCTTTTGTGTTTAAAGAGAA 7599

```

[illegible]

```

148 isLeuSerIleAspHisPheGlyValHisSerProAsnGluSerThrLeu 164
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
422 CTGAGTCTCCAGAAATCTAGGAGTCGAGCTTTAGATAAGCGTCATCTC 471
165 ValValThrLeuGluSerProThrSerHisPheLeuGluLeuAlaLeu 181
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
472 GAAATTCAGTTAGAACTCCCTCGCGGCAATTTCTACATTTCTTGACTCT 521
181 uProValPhePheProValHisLysSerGlnArgThrLeuGlnSer. .... 196
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
522 TCTATTTTTCCTGTTTCATGAACCTCTCGAACTATGACCACTCTT 571
197 ..LysLeuProIleAlaSerGlyAlaPheTyProLysAsnIleLys 212
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
572 TTGAAGAGATGCCCATTTACTCGCTGCTTTCCGCCCTGTCTCTAGAA 621
213 GlnLysGlnTrpIleLysLeuSerLysAsnProHisTyTrpAsnGlnSe 229
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
622 AAAGCCTGAGATCCATCTAGAGAAACCCCTATGTACCATATAAAG 671
229 rGlnValGluThrLysThrIleThrIleHisPheIleProAspAlaAsn 246
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
672 CCGTGTGAACACTACATAAAATATTGTACAGTTTATCTCAACCGCTAACA 721
246 hrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyProPro 262
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
722 CTGACGACCATCTATTCAAACTAAGAAATAGATTGGCAAGGACCTCT 771
263 TrpGlyGluArgIleProGlnGluThrLeuSerAsnGlnSerLysG1 279
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
772 TGGGAGAACTATCCCTCCAGAAATCTCAGCTTCTACATCAGATGA 821
279 yHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPheAsnI 296
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
822 CCAGCTCTTTTCTCTCCGGCGCTTCGACTACATAGTTACTCTTAATA 871
296 leAsnLysPheProLeuAsnMetLysLeuArgGluAlaLeuAlaLaser 312
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
872 TACAAAAAAACCTTCGACAACTGCTAAATTAATCCAGGCACTGAGCCT 921
313 AlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArgAlaLy 329
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
922 GCATAGACAAAGATATGTTAACCAAAAGTGTATACCAAGTCTTGCAGA 971
329 sThrAlaAspHisLeuLeuProThrAsnIleHis. .... SerTyProG 344
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
972 ACCTACAGATCATCTCTACATCCAAAGACTTTATCCAGGACCTATCCG 1021
344 luHisGlnLysGlnGluMetAlaGlnArgGlnAlaLysLysLysLeu 360
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1022 AACGGAAGACAAAC. .... GAAAGAATTTCTGAGGCTCAACAACTC 1065
361 PheLysGluAlaLeuGluGluGlnIleThrAlaLysAspLeuGluHi 377
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1066 TTTGAAGAAGCTCTAGACGAACCTTCAAAATGACACGCGAAGATCTAGAAA 1115
377 sLeuAsnLeuIlePheProValSerSerAlaSerSerLeuLeuValG 394
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1116 GGAACCTTGACTTCTCAACCTTTCTTTCTTACGGAAGATTGGC 1165
394 lnLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAlaIleProIle 410
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1166 AATGCTAAGACAACTAGGAAGAACTTAAATTTACTATCCCTATA 1215
411 ValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerGlyAsnPh 427
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1216 GTAGGCCAAGAGTTTTCACAAACAAAAAACTTCTAGAGGGGAACTA 1265
427 eSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAlap 444
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1266 TTCCCAACCGTGAACCAATGACCGCAGCATTTATTGATCCGATGCTT 1315
444 heLeuThrIlePheAlaTyProSerGlyValProProTyAlaIleAsn 460

```

```

: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1316 ATCTCATGATCTTCCCAATCTGGAGGAATTTCCCTATCACTCCAA 1365

461 HisLysAspPheLeuGluIleLeuGlnAsnIleGluGlnGluAspHi 477
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1366 GATTCACTCTTCAAACTCTCTCATAAAGATCACTCAAGAACATAAAA 1415
477 sGlnLysArgSerGluLeuValSerGlnAlaSerLeuTyTrpLeuGluThrp 494
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1416 ACACCTAGCAATCAGCTTATTATTGAAGCCCTTGACTATTATTGAACACT 1465
494 heHisIleIleGluProIleTyHisAspAlaPheGlnPheAlaMetAsn 510
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1466 GTCACATCTCGAACCACTATGTCATCCAAATCTTCGAATTCGTTGAAAC 1515
511 LysLysLeuSerAsnLeuGlyValSerProThrGlyValValAspPheAr 527
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1516 AAAACATTAATAAACTTTAATCTTTTGTTCGACGAACCTTCAGACTTTCG 1565
527 gTyAlaLysGlu 531
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1566 TTTTATAGAAAAA 1578

seq_name: gb_ba:AE002215
seq_documentation_block:
LOCUS AE002215 12980 bp DNA linear BCT 30-MAY-2000
DEFINITION Chlamydomophila pneumoniae AR39, section 43 of the complete
genome.
ACCESSION AE002215 AE002161
VERSION AE002215.2 GI:8163458
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 12980)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
JOURNAL
MEDLINE
20150255
PUBMED
10684935
REFERENCE
AUTHORS
2 (bases 1 to 12980)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7189472.
FEATURES
Location/Qualifiers
1..12980
/organism="Chlamydomophila pneumoniae AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(197..1609)
/genes="CP0560"
complement(197..1609)
/genes="CP0560"
/notes="sodium:sulfate symporter family protein; identified
by match to PFAM protein family HMM PF00939"
/codon_start=1
/transl_table=11
/product="sodium:sulfate symporter family protein"
/protein_id="AAF73687.1"
/db_xref="GI:8163459"

```



```

/translation="MNKKRFLSLFLTAVALLGWTFSPHSPASTNSNAWQFAIFTTIT
MGIIQFVPMGAIAGISLTLTQTLLEQGLSGFNPILANLWFLSFIAGKIITKG
LGERIAYFSAKGLSGLVITDFFLAPAIPSVTARAGGILYVPVTSLSDFG
SAAEQDGLISGLIKVAYSSVITSAWFLTAGNPLVAALAGHGVLSWILWAK
AAIPGLSLFLMPLIILYKLYPPKITSCEEARIRSAKLRKMGPKKEKTILMIFEL
LVLVTFDGLIGISATAALIGISLLILTNLDQKQVIAN"TAWETFFWFCALLMMA
SPNLGPIPLVAGDSAAALYSGLSWKGIFPLFLIYPSYHLFASNTAHIGAMYPFL
AVISLGNPTFAALATLASNLFGGLTHYSGFAPLYFGSHLTVVOEWRSGFALSI
NVIVIGISLWKKALGLI"
/translation="MMLDISHILEDLAYDEGILPREIAEAAIVKQMIIPTPYLLHLD
ATQRPVEIVNDGYQGHLYAMYLIAQFRESALPLIILKLFATEDDTPHAIAGDVLED
LPRILASVNDSDSLIKELIETKINPYVKAASISGLVTLVAGAGKIPRDKVIRYFAELL
NYRLEKOPSAWNLITAGICTLVPGLEYPIKSAFDGGLVDVTSFISWEDVENIHEET
VESCITHLCSSTELINDYLEEMKXWLEDFPIEP"
/translation="MNFLLNIKDLITITNPKRTLIENLSQLKENRNALVGEESG
GKTIITKALGLPENCILIKTSGILFEDIDITKLSPELHRIRGOKIATLQAMGSL
TPSMRTGMOIETLRQHKNKKEAYNAQQLLTDVCPNPKYSFQSYPELGGNRQ
RVVIALASQPKLILADEPTALDMSQAQVLRILRNKQOQKQATILLVTHNLSLVK
ELNCIDIITKDGKLIETGTVEEIFLSPKHPYTLKLLNAVSKLPIKKTSSPILANKFQP
LANSQOGL"
/translation="MENLSSAPSRSTWKSIIQNMVLGLTTLIILMLGALLLPWFYQ
DYEOTSLKDILVSPCSRFPFGDTLGRMFARTLGLSLIATLTIATLIDVCGLLW
ATVAISGKKIDFLMKRTIELRSLPFIILLIIPHGLPLIATITGWIPIIS
RIIYQGLLLKKNKPFVLSAKAMHASTFHILKHLPLNTLAPITISTIFTIPNAIYTEA
FISFLGLGIQPPQASLGTLYKEGINAIDYYPWLFPPSLMIALSISFNLIGEGAKTL
CLEEGSHG"
/translation="MFSYIKNRLLFNLLSLWTLVTLFVLVMTKTIIPGDPNDGQNVLS
EVLQTLKRSYGLDKPLYQOYQYQLHSTAKLDGNSLVYKDRKVTNIISTAPPSAIL
GLQSLFISIGGIALGTIAALKKKORSYILGASILQISIPAFITLLOQYVFAKIP
LLPTACWSSSTHILPTILALAVTPMAFIQUTYSVSAALNKDQVILLATAGLSPLKV
VHKHLPYALFPTISAFLLTITVDIPFAIENIFCIPGLGKWFICISIKQRDYPVALG
LSVYFGLFMLSLLSDLIQSIIDPQIRYAHGKRRK"
/translation="MNTVTFSTLQKSFLILEKLDSEFFFGTRTQILVITPTNIR
LAAKRGCKVSIKILILSFILLPVITAFILRFLHKRDKQFLCIPKVISNEDE
ALLGSRQAVAKREISPAFFSIPRYQIRIDTPKDDAPSILFPIGIEIILKDCI
DTLQNLKLRMDPLGHPEEKALFDSICEKQDQWMSLESKLLITHFLKYLFS
GIEQLNPGFNPNGRGYFSEISTAKIFHQHGRGYGRSSGPIKKEI"
/translation="MNGYQIFTRLKKGISFSSYNPNPNPFDEGCFYVWESQKSA
LDHGILLOKQTEFFYRN"
/translation="MLKFFLVCEERLCILTVATHRALLETPLAISFEKELKYYVRK
DILQHNKGFITLNTSPICS"
/translation="MGYQIFTRLKKGISFSSYNPNPNPFDEGCFYVWESQKSA
LDHGILLOKQTEFFYRN"
/translation="MNTVTFSTLQKSFLILEKLDSEFFFGTRTQILVITPTNIR
LAAKRGCKVSIKILILSFILLPVITAFILRFLHKRDKQFLCIPKVISNEDE
ALLGSRQAVAKREISPAFFSIPRYQIRIDTPKDDAPSILFPIGIEIILKDCI
DTLQNLKLRMDPLGHPEEKALFDSICEKQDQWMSLESKLLITHFLKYLFS
GIEQLNPGFNPNGRGYFSEISTAKIFHQHGRGYGRSSGPIKKEI"

```

```

/translation="MNKKRFLSLFLTAVALLGWTFSPHSPASTNSNAWQFAIFTTIT
MGIIQFVPMGAIAGISLTLTQTLLEQGLSGFNPILANLWFLSFIAGKIITKG
LGERIAYFSAKGLSGLVITDFFLAPAIPSVTARAGGILYVPVTSLSDFG
SAAEQDGLISGLIKVAYSSVITSAWFLTAGNPLVAALAGHGVLSWILWAK
AAIPGLSLFLMPLIILYKLYPPKITSCEEARIRSAKLRKMGPKKEKTILMIFEL
LVLVTFDGLIGISATAALIGISLLILTNLDQKQVIAN"TAWETFFWFCALLMMA
SPNLGPIPLVAGDSAAALYSGLSWKGIFPLFLIYPSYHLFASNTAHIGAMYPFL
AVISLGNPTFAALATLASNLFGGLTHYSGFAPLYFGSHLTVVOEWRSGFALSI
NVIVIGISLWKKALGLI"
/translation="MMLDISHILEDLAYDEGILPREIAEAAIVKQMIIPTPYLLHLD
ATQRPVEIVNDGYQGHLYAMYLIAQFRESALPLIILKLFATEDDTPHAIAGDVLED
LPRILASVNDSDSLIKELIETKINPYVKAASISGLVTLVAGAGKIPRDKVIRYFAELL
NYRLEKOPSAWNLITAGICTLVPGLEYPIKSAFDGGLVDVTSFISWEDVENIHEET
VESCITHLCSSTELINDYLEEMKXWLEDFPIEP"
/translation="MNFLLNIKDLITITNPKRTLIENLSQLKENRNALVGEESG
GKTIITKALGLPENCILIKTSGILFEDIDITKLSPELHRIRGOKIATLQAMGSL
TPSMRTGMOIETLRQHKNKKEAYNAQQLLTDVCPNPKYSFQSYPELGGNRQ
RVVIALASQPKLILADEPTALDMSQAQVLRILRNKQOQKQATILLVTHNLSLVK
ELNCIDIITKDGKLIETGTVEEIFLSPKHPYTLKLLNAVSKLPIKKTSSPILANKFQP
LANSQOGL"
/translation="MENLSSAPSRSTWKSIIQNMVLGLTTLIILMLGALLLPWFYQ
DYEOTSLKDILVSPCSRFPFGDTLGRMFARTLGLSLIATLTIATLIDVCGLLW
ATVAISGKKIDFLMKRTIELRSLPFIILLIIPHGLPLIATITGWIPIIS
RIIYQGLLLKKNKPFVLSAKAMHASTFHILKHLPLNTLAPITISTIFTIPNAIYTEA
FISFLGLGIQPPQASLGTLYKEGINAIDYYPWLFPPSLMIALSISFNLIGEGAKTL
CLEEGSHG"
/translation="MFSYIKNRLLFNLLSLWTLVTLFVLVMTKTIIPGDPNDGQNVLS
EVLQTLKRSYGLDKPLYQOYQYQLHSTAKLDGNSLVYKDRKVTNIISTAPPSAIL
GLQSLFISIGGIALGTIAALKKKORSYILGASILQISIPAFITLLOQYVFAKIP
LLPTACWSSSTHILPTILALAVTPMAFIQUTYSVSAALNKDQVILLATAGLSPLKV
VHKHLPYALFPTISAFLLTITVDIPFAIENIFCIPGLGKWFICISIKQRDYPVALG
LSVYFGLFMLSLLSDLIQSIIDPQIRYAHGKRRK"
/translation="MNTVTFSTLQKSFLILEKLDSEFFFGTRTQILVITPTNIR
LAAKRGCKVSIKILILSFILLPVITAFILRFLHKRDKQFLCIPKVISNEDE
ALLGSRQAVAKREISPAFFSIPRYQIRIDTPKDDAPSILFPIGIEIILKDCI
DTLQNLKLRMDPLGHPEEKALFDSICEKQDQWMSLESKLLITHFLKYLFS
GIEQLNPGFNPNGRGYFSEISTAKIFHQHGRGYGRSSGPIKKEI"

```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/translation="MKMRLKPTLKLIPNLLFLITLSSCKOKOPLKGLHVIAMS
HDLADPNNAVLYSDKSAKALYEGTRETDOGLALAESEYTLSDKHVYTFKLRP
SVMSDCTPUTAYDEFSKISOLYFEFSPSIHTLIGVINKNSAIHNAOKSLETIGIOAK
DDLTLVLTLEOPPYELTLARPVSPVHHTLDESITKGTTPSYISNGPFVLKKEH
QNVLLIKNPHYIHESYKDRVTLKIIPDASTATKLFKSIDWIGSPWSAPISNEH
QKVLSEKILYISVSTLLIYNLQKPLONKALRAKHAIDRKSILRLVPSGOEAV
TLVPPNLSQNLQKEIESTEERQTKARAYFOEAKETLSEKELAEISILYPTDSSNSII

alignment_scores:

Quality: 1160.50 Length: 521
Ratio: 2.923 Gaps: 4
Percent Similarity: 76.200 Percent Identity: 43.186

alignment_block:

US-09-824-567-2 x AB002215/rev ..

Align seg 1/1 to reverse of: AB002215 from: 1 to: 12980

15 SerLeuSerValValLeuGlnGlyCysLysGluSerSerHisSerSerth 31
12848 rCTCTTTGGCTCATCTACTCAGGATCT.....TCCCATCCCAATC 12808
31 rSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuA 48
12807 CTCCTAAAGGAATTTTGTGTAATATGAGGAATGCCACCTCTCTGG 12758
48 sProArgGlnValArgLeuSerGluIleSerLeuValLysHisIle 64
12757 ATCTCGAANAACCTGCTCTCATTCGACACCAACTCTAATCGCTCATCTA 12708
65 TyrGluGlyLeuValGlnGluAsnLeuSerGlyAsnIleGluProAl 81
12707 TATGAGACACTGCTCGAAGAACATTCCTCCAAAATGAGAGATTAACCCAGC 12658
81 leuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheI 98
12657 CCTTCAGAAAGCTACACCATCTCCGACAGCGGACTCGGTACACATTTA 12608
98 ysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAsp 114
12607 NAATCAAAACATCTTTGAGTAACGGAGACCCCTCGACAGCTCAAGAC 12558
115 PheIleGluSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTy 131
12557 TTTGCTCTCTTGGAGGAATCTTAAGGAAGATCGCTCCCTCGGTATA 12508
131 rAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGlnGluLys 148
12507 TCTCTATCGCTTTTACCTATCAAAATGCTCGGCAATCTTTGATGATA 12458
148 lsLeuSerIleAspHisPheGlyValHisSerProAsnGluSerThrLeu 164
12457 CTGAGTCTCCAGAAATCTAGAGTCCGAGCTTTAGATAAGCGTCACTTC 12408
165 ValValThrLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaI 181
12407 GAATTCAGTTAGAACTCCCTCGCGGATTTCTACATTTCTTGACTCT 12358
181 uProValPhePheProValHisLysSerGlnArgThrLeuGlnSer.... 196
12357 TCCTATTTTTCCTGTTCATGAACTCTCGAAGACTATACACCTCTT 12308
197 ..LysSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIleLys 212
12307 TTGAAGAGATGCCCATTTACTCGCGTGTCTTCCGCCCTGTGCTCTAGAA 12258
213 GlnLysGlnTrpIleLysLeuSerLysAsnProHisTyrTyrAsnGlnSe 229
12257 AAGGCTGAGACTCCTCTAGAGAAACCCCTATGTACCAATAATPAAAG 12208
229 rGlnValGluThrLysThrIleThrIleHisPheIleProAspAlaAsnT 246
12207 CCGTGTGAACACTACATAAAATATTGTACAGTTTATCTCAACCGCTAACA 12158

246 hrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyProPro 262
12157 CTGAGCCATCTTATTCAACAATAAGAAATAGATTGGCAAGGACCTCT 12108
263 TrpGlyGluArgIleProGlnGluThrLeuSerAsnLeuGlnSerLysG 279
12107 TGGGAGAACCTATCCCTCCAGAAATCTCAGCTTCTCTACATCAAGATGA 12058
279 yHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPheAsnI 296
12057 CCAGCTCTTTTCTTTCGGCGCTTCGACTACATGTTACTCTTTAATA 12008
296 leAsnLysPheProLeuAsnAsnMetLysLeuArgGluAlaLeuAlaSer 312
12007 TACAAAAAACCCTTGGACAACTGCTAAATACGCAAGGCTTGGAGCCTT 11958
313 AlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArgAlaL 329
11957 GCAATAGACAAAGATATGTTACCAAGAGTGGTATACCAAGGCTTGCAGA 11908
329 sThrAlaAspHisLeuLeuProThrAsnIleHis.....SerTyrProG 344
11907 ACCTACAGATCATCTCTACATCCAGACTTTATCCAGGACCTATCCCG 11858
344 luHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeu 360
11857 ACGGAAAACAGCAAAAC.....GAAAGAAATTTTGGAGGCTCAACAAC 11814
361 PheLysGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGluH 377
11813 TTTCAAGAGCTCTAGACGACTTCAATATGACACCGGAGATCTAGAAA 11764
377 sLeuAsnLeuIlePheProValSerSerAlaSerSerLeuLeuValG 394
11763 GSAACCTTTGACTTTCTCACTTTCTTTTCTTACGGAAGATTTGCC 11714
394 lnLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAlaIleProIle 410
11713 AAATGCTAGAGAACATGGAAGAAAGTCTTAAATTTACTATCCCTATA 11664
411 ValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPh 427
11663 GTAGGCCAAGAGTTTTCACAAATCAAAAAAACTTCTTAGAGGGGAAC 11614
427 eSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAla 444
11613 TTCCCTAACCTGACCAATGACCGACGATTTATTGATCCGATGCTT 11564
444 heLeuThrIlePheAlaTyrProSerGlyValProProTyrAlaIleAsn 460
11563 ATCTCATGATCTTTGCCAATCTCTGGAGAAATTTCCCTATCAGCTCCAA 11514
461 HisLysAspPheLeuGluIleLeuGlnAsnIleGluGlnGlnAspHI 477
11513 GATTCACACTTTCAAACTCTCTCATATAAGATCACTCAAGACACATAAAA 11464
477 sGlnLysArgSerGluLeuValSerGlnAlaSerLeuTyrLeuGluThrP 494
11463 ACACCTACGAAATCAGCTTATTATTGAGGCCCTTGACTATTATTAGAAC 11414
494 heHisIleIleGluProIleTyrHisAspAlaPheGlnPheAlaMetAsn 510
11413 GTCACTTCTCGAACCACTATGTCATCCAAATCTCGAATTTGCTTTGAAC 11364
511 LysLysLeuSerAsnLeuGlyValSerProThrGlyValValAspPheAr 527
11363 AAAACATTAANAACCTTAACTTTTGTTCGACGAACCTTCAGACTTTCC 11314
527 gTyrAlaLysGlu 531
11313 TTTTATAGAAAAA 11301

[illegible]

199 ACCCTTAGGAAA...CATCTCGTTATTCGATGAGCCATGATCTCGCCG 245

```

342 yrProGluHisGlnLysGlnGluMetalaGlnhrgGlnAlaTyrAlaLys 358
1131 TCAATCTTCAAAAGAGATCTCACAGAGAACGACAAACAAAGCCAGA 1180
359 LysLeuPheLysGluAlaLeuGluGlnGlnIleThrAlaLysAspLe 375
1181 GCATATTTTCAAGAGCTAAAGAACACTT.....TCTGAAAGAACT 1224
375 uGluHisLeuAsnLeuIlePheProValSerSerSeraLeuSerLeu 392
1225 CGCAGAACTCAGACTCTCTATCTATAGATTCCTCGAATTCCTCATCA 1274
392 euValGlnLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAla 408
1275 TAGCTCAAGAACTCAAGAGACTTAAAGATACCTTAGGATGAAATC 1324
409 ProIleValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerG 425
1325 AAAATCAAGGATGAGTACCCTGCTTTTAAAGAAACGTCGTCRAGG 1374
425 YAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAsp 442
1375 AGATTTCTCATAGCAGCAGAGAGTATCGGAATACGTAAGCCCG 1424
442 etAlaPheLeuThrIlePheAlaTyrProSerGlyValProPTrAla 458
1425 TAGCCTTCTATCTATTCTAGCAACCCAGAGACTCTACACATGG... 1471
459 IleAsnHisLysAspPheLeuGluIleLeuGlnAsnIle.....G 472
1472 ...AGAAACAGTGATTCAGAAAGACTTTAGAGAACTCTATCTCCCTCA 1518
472 uGlnGluGlnAspHisGlnLysArgSerGluLeuValSerGlnAla 489
1519 TGCTTACAAAGAGATTTAAACCGCAGAAATGATA..... 1555
489 euTyrLeuGluThrPheHisIleIleGluProIleTyrHisAspAla 505
1556 ....ATAGAAGAAAGAACCCGATTCCTCCCTGTATCAGCGCAATAT 1600
506 GlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProTh 522
1601 ATTACGCTATACATCTCTAAATCCAGATATCCGATCTCTCTAGG 1650
522 yValValAspPheArg 527
1651 CCACACAGATCTCAA 1666
seq_name: gb_ba:AE002313
seq_documentation_block:
LOCUS AE002313
DEFINITION Chlamydia muridarum, section 44 of 85 of the complete genome.
ACCESSION AE002313
VERSION AE002313.1
KEYWORDS
SOURCE
ORGANISM Chlamydia muridarum.
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
1 (bases 1 to 11402)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Bowman,C., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
McClarty,G., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
Medical Center Dr, Rockville, MD 20850, USA
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
20150255
PUBMED 10684935
2 (bases 1 to 11402)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,

```

```

Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1..11402
/organism="Chlamydia muridarum"
/db_xref="taxon:83360"
/feature="synonym: Chlamydia trachomatis MoPn"
250..1569
/gene="TC0440"
250..1569
/gene="TC0440"
/feature="similar to GP:2444077; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phospholipase D family protein"
/protein_id="AAF39294.1"
/db_xref="GI:7190486"
/translation="MCSPCPRLHPIVTDCTSTPLRRPLSSTTAAVVVVQLFK
YHRKLSHRVIYDYGKNTKPKIQSSNRHSLATSSKTSRKLIRARTRTSVVP
RPIAQQLITFIQAQKTSNPLKICKFIDRTQNLHIQVCFNHPTIIQHLVAAA
REVSVYVRDGTLEAEKSTICLOKSGRSILPHKSMGVDSORVSLASNGFTPOS
TEODINLSLINDSELARHETNOFTVQIGNQFLYIPIFRNKDRLAVRTIKCI
AQAKTIVVVVILLHPVLKSLQAAAARGVKVEAVDTRSEOTORTLERLQSLPL
AVKPGAPRVHAKCYIDNKILICGSANSSAGLTNRREDLFIIRGLTETQORLSFI
WQSVKEATPLTAQSLKPRFEEDDPGEGTSSGISAGASAKAKTQ"
/gene="TC0441"
complement(1661..3049)
/gene="TC0441"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39295.1"
/db_xref="GI:7190486"
/translation="MVDKFFFIQQLSLLTSELSDTPTQDTKGLPKADIVHLP
GTISPEAWELGVNRGLIKWKNQWSPAFSGNPHKSVAEIRFNFSIRYERPNVS
LLKAITTRDFASDFRIMATVQGHPPGIGIONDLWLQNLVLCIQDNIILYEV
QQNRIAYVYPSLPLEARLYDLFASQLFLSOGITLFLNCFNKNGSSNLOQS
TQORSEAAWLDRAASFNNLVFVLSAGASCEPAATKLTSGYHATENGFGCEA
HAGETGFLYLNQITQALPIORAHGFOATEHVPTIHAIOEKNTLVMAFAINVLGA
SLHOSGLEKISKTMINLDEHPFALFRDHLKLSVLSNDPQMGTSVQNTMLLSG
FSADESHLTHVTSSPTTFKEIQLNVEAIVSFVDVDTKIMLLNSIHEVLTSLK
KQYNLLEGRKFL"
complement(3064..4602)
/gene="TC0442"
complement(3064..4602)
/feature="similar to GB:M10101 SP:P04079 PID:146276 GB:U00096
PID:1788854; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="GMP synthase"
/protein_id="AAF39296.1"
/db_xref="GI:7190487"
/translation="MSNIIIDFGSOYTNVLAKKIRLLSVFCEVLPWNTPEKILOKS
PSGLIFSGSHSVVQNNSPEDVKEIYNSNIPILGVCQMQLIARDGSEVQGGKSEFG
YPIVYVSELFKGLADQADAFHEIRMSHCDVVPKDFVIASSOHCPIAIBCPK
KLFLGQHPVEVSSQDQIGKILNSNHIQTSITKTIETKLIQIGIKVGETE
RVLLSGGVDSVYAVLLHNLALRSLCVFDTGLLRKNEVEEVKQFSSLSGELV
VDAEQTFHDLKSGIEDPEQKRKVGAAFEVFEASKNLDVQWLAQCTIYSDVIESAK
SCDAQTVIKSHNVGGLEPEKLNKLEPLFLFDEVKGLVGLPDLVLSRHPFG
PGLGVRLGVEVRREYVEIVKNSDIFIELKANLIHKVSOAFVLPCKVAVKGDG
RHYGTIALRAVESTDFWACWPSUSREPLNCSRIINEIPEVCRVYDIDSKPPAT
LSWE"
complement(4618..5691)
/gene="TC0443"

```



```

7680 TGGCTTACGATTGCTTGGCTTAATAATAAATTCATGCTGTTCTAAAG 7729
147 lyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSerThr 163
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7730 GAGACATTCCTGTTGAAATTTAGTGTTCGCGCTTTAGATGAACATACT 7779
164 LeuValValThrLeuGluSerProThrSerHisPheLeuLeuLeuAla 180
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7780 CTTGAAATCACTCTAGAACCCCTCTTCATATTTTAGAATCTCTAC 7829
180 aLeuProValPhePheProValHisLysSer.....GlnArgT 193
: ||||| : : : : : : : : : : : : : : : : : : : : : : :
7830 ACATCCCGTTTTTATCCAGTCAATGCTCTTTGAGAGAAATATATAGAA 7879
193 hrLeuGlnSerLysSerLeuProIleAlaSer...GlyAlaPheTyPro 208
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7880 ATCGAAGTAAACGCTCTCCCAATATTTCTAACGGCGCTTTTATCAT 7929
209 LysAsnIleGlyGlnLysGlnTrpIleLysLeuSerLysAsnProHisTy 225
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7930 CGTGTTATGAGCCACAAAATTTTGTGCTTGATATAAAATCCTTTTA 7979
225 rTyAsnGlnSerGlnValGluThrLysThrIleThrIleHisPheIleP 242
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7980 TCATGCCAAAATATGTTCTTTAGATGCTGTACGCTTACAAATAGTTC 8029
242 roAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrp 258
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8030 CCGATATTCATACAGCGTGCATTTGTTCAAAAAAATATGTCGATTTA 8079
259 GlnGlyProProTrpGlyGluArgIleProGlnGluThrLeuSerAsnLe 275
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8080 GTGGGTTGCTTGGAGCTCTCTCCCTCCAGAGAACAAAATATCT 8129
275 uGlnSerLysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpL 292
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8130 T...TCTCAAGATTTTGTATGATTATCCGGTGCTAAATGCACAGTTC 8176
292 eThrPheAsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGlu 308
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8177 TATTCGTAATGTTAACCAACACCATAGATATCTTCGTAAGAGCG 8226
309 AlaLeuAlaSerAlaLeuAspLysGlnAlaLeuValSerThrIlePheLe 325
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8227 GCATCTCTTTAGCCATTTGATCGAGAAACATATTAAAA.....TTAC 8270
325 uGlyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSerT 342
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8271 TGGCAAGGAGGATGCTACCAAGTTTGTATCATCTCTAGCTATAAGA 8320
342 yrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyAlaLys 358
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8321 TGCCTTTAGATGTTCTTTCGCAA...AAGAGAGAAATCTCTTAGCTAAA 8367
359 LysLeuPheLysGluAlaLeuGluLeuGlnIleThrAlaLysAspLe 375
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8368 AATTACCTACGGAACCTTTAAA.....ACTGTCCTCAAGAGAGTT 8411
375 uGluHisLeuAsnLeuIlePheProValSerSerSerAlaSerSerLeuL 392
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8412 AAAAAAATCACTACTCTACCCATAGATATCTATTGCTTCGAGCAG 8461
392 euValGlnLeuArgGluGlnTrpLysGluSerLeuGlyPheAlaIle 408
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8462 TTGTTCAAGAGATCTCTCAACAGTATTGTATGTTTAGGATTTAGATT 8511
409 ProIleValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerG 425
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8512 TCTACTTTAGGATTAAGATACCAAGATTTTTCGATAAAACGTTCTAAAG 8561
425 yAsnPheSerLeuAlaTrpGlyTyPheAlaAspPheAlaAspProm 442
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8562 AGAGTTTCTTTATCAACAGGAATTCGGTGTCTGATTATCAACAGCCA 8611

```

```

442 etAlaPheLeuThrIlePheAlaTyProSerGlyValProProTyAla 458
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8612 AAGCTTTCTATCCATCCCTGTGATGGAGCTAAATATAAAGATTATCAA 8661
459 Ile.....AsnHisLysAspPheLeuGluIleLeuGlnAsn..... 470
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8662 GTGATTTGATTGGCAAAATCAAGAATATACGGATATCGTATCTCGACTTT 8711
471 IleGluGlnGluGlnAspHisGlnLysArgSerGluLeuValSerGlnA 487
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8712 GTTGAGGATTCAACAGATTTCACAAATTTCTAGCAGAA..... 8748
487 laSerLeuTyLeuGluThrPheHisIleIleGluProIleTyHisasp 503
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8749 ..CAACTATTGCTCAAGAGATCCCTCTTAT...CCCTTATACACTTG 8793
504 AlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerPr 520
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8794 GATTATCTTACGCTAAACATCTAAAGTCTCTAATCTCAAAACCTCCTC 8843
520 oThrGlyValValAspPheArg 527
||| : : : : : : : : : : : : : : : : : : : : : : : : :
8844 ACTCGAGAAATTTGATTAAAA 8865

seq_name: gb_ba:AE001291
seq_documentation block:
LOCUS AE001291 10827 bp DNA linear BCT 30-OCT-2000
DEFINITION Chlamydia trachomatis section 18 of 87 of the complete genome.
ACCESSION AE001291 AE001273
VERSION AE001291.1 GI:3328573
KEYWORDS
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1 (bases 1 to 10827)
Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
Science 282 (5389), 754-759 (1998)
PUBMED 9784136
REFERENCE
2 (bases 1 to 10827)
Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)
99206606
PUBMED 10192388
REFERENCE
3 (bases 1 to 10827)
Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Direct Submission
Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
source
1. .10827
/organism="Chlamydia trachomatis"
/strain="D/UW-3/Cx"
/db_xref="taxon:813"
206. .1384
/gene="trpS"
/note="CT170"
206. .1384
/gene="trpB"
/codon_start=1
/trans_table=11
/product="tryptophan Synthase (Beta Chain)"
/protein_id="AAC67761.1"

```

```

/db_xref="GI:3328574"
/translation="MFKKHPFGAFLPBELLAPIONLKAWEILKTOQSFSLSELDGI
LNKYAQRTPLTEVKNFPAIDGRVFLKDELLHTGAHKLNNALGCLLAKYLKTR
CGVAGDVAQNALQWATATACAYLGLDCVYKACDVEROKPNVEKMRGLAEVYVSTIGR
VCLGTQAGNOLADWATATSFHYHCLGSLGLPLPYDPIVRFQSVISAEVKQIIRHAGV
RQPDILLACGGGSNAIGFFHHFIPNPKVOLLIGPVGGLGISGKHAAFPATGPGVF
HGFSYLLQDDQQLYQIHTSISAGLDPSVGDPAHMHESGRAFTLATDEKLAARFF
LLTRNEGTLIAPLESSHAHLVSIAPSLKEQIVIVNLGRGDKDLPLQIIRNRNGIVE
1377..2138
/gene
/gene="trpA"
/note="Ct171"
1377..2138
/gene="trpA"
/codon_start=1
/transl_table=11
/product="Tryptophan Synthase (Alpha Chain)"
/protein_id="AAC67762.1"
/db_xref="GI:3328575"
/translation="MSKLTQVFKTKLIGVLTAGDGGSTYIEAAKALIQGVWDILE
LGFSPDPVADNPETQVSHDRALAEMLTSEILLEIVGRAFNQEVLIIYSYNNLL
QRDLIDRLKLDAGINGVCVIDLPAFLSHGSKSPFFEDLLAVGLDPLILLSAGTTPER
MSLQETARGFLLYIPCAQTRDSEVGIKEEFKRVREHFDLPIDVRRDIOCKKEAAHVL
NLSQGIYATFVFWHQTMTDMSVETLTAAQTVPG"
/complement(2385..2876)
/gene="Ct172"
/complement(2385..2876)
/gene="Ct173"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67763.1"
/db_xref="GI:3328576"
/translation="MNYHNTFVKTSMFLAKRIVOLKNKPPILLKFKSETVLEIFPERQ
LKMWEGYSIDENNYTSDYNNFERGPILOKLANPVCALLOKLOEQAQMTLSNQVTVG
DIVLMRSPIFKFSVLLETLINELIQESLFLFAKPNVOCPKNSFEHGAHEILLKLTIFL
TVS"
/complement(2904..3074)
/gene="Ct172.1"
/complement(2904..3074)
/gene="Ct172.1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67773.1"
/db_xref="GI:3522888"
/translation="MLRLFCGYFFQLFEKQKNLLEDLYNKLVFETVVMGEGACVV
NLAIGGETILL"
/complement(3142..3414)
/gene="Ct173"
/complement(3142..3414)
/gene="Ct173"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67764.1"
/db_xref="GI:3328577"
/translation="MKICREICDILKRDFAFTEIERTYKNGSIVYTSRQCPGVFEIR
IGDQAKLHMQEALQAVATLGLIKHLIIPPLSIYKHSLLYKRVLT"
/complement(3501..3956)
/gene="Ct174"
/complement(3501..3956)
/gene="Ct174"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67765.1"
/db_xref="GI:3328578"
/translation="MLIFEDKSQSGALPDLREAGNLIRFAVNRGMASQIKVTSQSG
HIFTSEKMSIVCRACIYVLCVLIAPFCILGALIGTIAYKLSYQNALYLFREHRNK
CSEVEMKMGKNKQITRLQRFNRFLVEKKHIAADVKKQEQEMCRQSES"
4151..5740
/gene="oppA_2"

```


5689 CTCCTCTCGGAGAAATTGATTAAAA 5716

seq_name: gb_ba:BACSP00K

seq_documentation_block:

LOCUS BACSP00K 5300 bp DNA linear BCT 06-MAR-1995
 DEFINITION Bacillus subtilis spo0K operon.
 ACCSSION M57689
 VERSION M57689.1 GI:143602
 KEYWORDS Spo0K operon; oligopeptide permease; sporulation protein.
 SOURCE B. subtilis (strain JH642) DNA, clones pDR20/21, pJL2/3 and pJL7.
 ORGANISM Bacillus subtilis
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus.

REFERENCE 1 (bases 1 to 6300)
 Rudner,D.Z., LeBeaux,J.R., Ireton,K. and Grossman,A.D.
 TITLE The spo0K locus of Bacillus subtilis is homologous to the
 oligopeptide permease locus and is required for sporulation and
 competence

JOURNAL J. Bacteriol. 173 (4), 1388-1398 (1991)
 MEDLINE 91139580

FEATURES
 source

Location/Qualifiers
 1..6300

/organism="Bacillus subtilis"

/strain="168"

/sub_strain="JH642"

/sub_species="Marburg"

/db_xref="taxon:1423"

/clone="pJL7"

/clone="pDR20/21"

/clone="pJL2/3"

477..2114

/gene="spo0KA"

477..2114

/gene="spo0KA"

/codon_start=1

/transl_table=11

/product="sporulation protein"

/protein_id="AAA62687.1"

/db_xref="GI:143603"

/translation="MKRWSIVTLMILFTLVLSAGFGSGSGEGKSKSGKTTNLN
 NKTEPFLHPLGANDSVSGVIRQTPEGLTRINADGPEEGMASKLETSGKNTTIF
 TIRGKVSNGSDPYTAQDFEYANKWALDPNNSQAYQLYIKGAEAAATGKSGLDDV
 AKVANDKTLKVLNNEPTVTELTAFYIMPINKKIAEKKNWNTNAGDDVYVNSGPF
 KWTAKSGSLTEKNDQWYDKVKKIKKIDVMNNNTLKKFQAGELDAGMPLG
 QPTESLTKKGSLSHVEPIAGVYKTEAKPLDNVIRKALYSLDRQSIYKVN
 TUGEQIPAMAAVPTMKGFEDNKEGVDKNDVYKAEVLEKGLKEMGLSKASDLPIK
 LSYTDDAHAKIAQVQEMWKNLGVVDLNSWNVYIDKLHSOQYQIGRWGLDF
 NDPINFLFDRKNGNDTGWENPEFKLLNQSTEDTKRAELLKRAEGIFIDEM
 PVAPIYFTDTWQDENLKGVIIMPGTGEVYFRNATFK"

sig_peptide

/gene="spo0KA"

537..2111

/gene="spo0KA"

/product="sporulation protein"

222..3157

/gene="spo0KB"

2222..3157

/gene="spo0KB"

/codon_start=1

/transl_table=11

/product="sporulation protein"

/protein_id="AAA62688.1"

/db_xref="GI:551726"

/translation="MLKYIGRLVYMIITLFIIVITVFFLMQAAPGPGSPGKLLPE
 TEANLNAHYGLDKPLFVQVYLSKVAMWDPGPKYKGSVNDLISGFPVSPFLGA
 BAILLALAGLVGVAALYHNKQWDTVALIIFGVSFIMAAYVQVFSMKLGL
 FPVAGWDSNAYTFLPSIALASMPMAFIARLSRSMIEVLNSYIETAKAGLSAQRLQ
 CGTFEFHRCRLHVLGPNMAAQLVGSFIETIFGIPGLGAHFVNSITNRDYTVIMVG
 TVFFSVILLCLVLDVLYGIIDPRIKLSKAKKA"

CDS

2255..3157

/gene="spo0KB"

/note="alternative translation start"

/codon_start=1
 /transl_table=11
 /product="sporulation protein"
 /protein_id="AAA62689.1"
 /db_xref="GI:143605"
 /translation="MIILFVIVTVFFLMQAAPGPGSPGKLLPEIEANLNAHYGL
 DAPLFVQVYLSKVAMWDPGPKYKGSVNDLISGFPVSPFLGAEAAILLALGV
 LFGVIAALYHNKQWDTVALIIFGVSFIMAAYVQVFSMKLGLFVAGWDSNAY
 TFLPSIALASMPMAFIARLSRSMIEVLNSYIETAKAGLSAQRLQCCQTPETHFQR
 LLHILGPNMAAQLVGSFIETIFGIPGLGAHFVNSITNRDYTVIMGVTVFFSVILLCL
 VLDVLYGIIDPRIKLSKAKKA"
 3161..4078
 /gene="spo0KC"
 3161..4078
 /gene="spo0KC"
 /codon_start=1
 /transl_table=11
 /product="sporulation protein"
 /protein_id="AAA62690.1"
 /db_xref="GI:143606"
 /translation="MQNIKPNMFEPAAANAGDAEKISKSLSLWKDAMLPPRSNKIAM
 VGLIIVIIILMAIFAPMFSDYDSTNLLNADKPPSKDHWGCTDGLGDFVPTWVG
 ARISIFIGVAAAVLDLLIGVWISGSGRGRTDEIMRTADILNAPSLMLVILMV
 VLPKGLFTIITAMITGWINMARIVRGVOLKNOEYVLSOTLGAKTSLRLFKHIVP
 NAMGSLVLTWLTVPITAFTEAFSLYSLGVFAPLASWGTMSDGLPALTYTPWRLLF
 PAGFICITMFGFNVGDLRDALDLPKLRK"
 4083..5159
 /gene="spo0KD"
 4083..5159
 /gene="spo0KD"
 /codon_start=1
 /transl_table=11
 /product="sporulation protein"
 /protein_id="AAA62691.1"
 /db_xref="GI:143607"
 /translation="MIRVTRILEVKDLAISFKTYGGEVQAIRGVNPHLDKGTETIAIVG
 ESSGSKVTSOAIMKLIIMPFGYKPKGEILFEGKDLVPLSEREMQNVKGIEMIFOD
 PMTSLNPTMKVKGQITEVLFKHEKISKRAAKRAVELLELVGIMPERRVNPQPFHS
 GMRQRVVAMALAANPKLLIADDEFTALDVTQAILKMDKLOKKTIDTSIIFTHD
 LGVVANVADRVAMVAGQIVETVDEIFYDPRHPTWGLLASMPTESSGEELTAI
 PGTPDITNPKPDAFALRSSYAMKIDFEQEPPEMKVSDTHYVKSLLHPAPKVEPP
 EAVKAKMRKLANTFEKPVIVREGE"
 5152..6078
 /gene="spo0KE"
 5152..6078
 /gene="spo0KE"
 /codon_start=1
 /transl_table=11
 /product="sporulation protein"
 /protein_id="AAA62692.1"
 /db_xref="GI:143608"
 /translation="MNETKLEIKHLKHQFVTPRGTVKAVDDLSFDIVKGTGLGV
 GESGCKSTTGRSIIIRLYEATDGVLPNGENVHGRKSKKLLFNKRMQMIPODPYAS
 LNPRTVADIIAEGDLIHLAKTKKERMORVHELLETVGLNKEHANYRPHFSGGQRQ
 RIGIARALAVDPEFTIADDEPISALDSIQAOVNVLMKELQKGLTYLFIADLSNVK
 YISDRIGVYMGKGLVELAPADELTENPLHPTKSLLSAIPLPDPDYERNVRQKDYPS
 VHOLKDGTEMEFRVKPGHFVYMCTEAEFKAFS"
 5161..6078
 /gene="spo0KE"
 /note="alternative translation start"
 /codon_start=1
 /transl_table=11
 /product="sporulation protein"
 /protein_id="AAA62693.1"
 /db_xref="GI:551727"
 /translation="MTEKLEIKHLKHQFVTPRGTVKAVDDLSFDIVKGTGLGVES
 GSKTTSGRSIIIRLYEATDGVLPNGENVHGRKSKKLLFNKRMQMIPODPYASINP
 RARLADITAEGLDHLAKTKKERMORVHELLETVGLNKEHANYRPHFSGGQRQIG
 IARALAVDPEFTIADDEPISALDSIQAOVNVLMKELQKGLTYLFIADLSNVKIS
 DRIGVYMGKGLVELAPADELTENPLHPTKSLLSAIPLPDPDYERNVRQKDYPSVHQ
 LADGETMEFRVKPGHFVYMCTEAEFKAFS"
 1919 a 1222 c 1462 g 1697 t
 BASE COUNT
 ORIGIN

seq_name: qb_pat:AX027646

[illegible]

```

1886 AGGCTGGGAAAAATCCAGAAATTCAGAAAGCTTCTGAATCAGTCACAAACTG 1935
474 luGlnAspHisGlnLysArgSerGluLeuValSerGlnAla...SerLeu 489
1936 AAACAGATAAAACAAACACGTCAGAGCTGCTCAAAAAAGCAGAGGTATT 1985
490 TyrLeuGluThrPheHisIleIleGluProIleTyrHisAspAlaPheG 506
1986 TTCATTGATGAATGCCG...GTTGCCCAATCTATTTCATATCATGATAC 2032
506 nPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProThrGlyV 523
2033 TTGGGTACAGGATCAAAACCTAAAGGTGTTATCATGCCAGGTACTGGTG 2082
523 alValAspPheArgTyrAla 529
2083 AGGTTTATTTCAGAAACGCA 2102
```

Sequence	Strid Orig	5Score	EScore Len	! Documentation
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20238	..	2739.00	4877.44	2.5e-263
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20190	..	2726.00	4779.32	7.3e-251
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX201425	..	1356.00	2326.35	3.1e-121
SIDS5/gcgdata/geneseq-genemb1/NA2002.DAT.AAS18750	..	755.50	1232.26	2.8e-65
SIDS5/gcgdata/geneseq-genemb1/NA2000.DAT.AA51864	..	617.50	1061.54	8.8e-51
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AA51764	..	505.50	877.02	1.7e-40
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AA51763	..	505.50	876.65	1.8e-40
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20248	..	505.50	803.83	2.0e-36
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20020	..	455.00	785.65	2.1e-35
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20021	..	455.00	763.22	3.7e-34
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20021	..	452.50	781.69	3.4e-35
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20253	..	451.50	747.48	2.7e-33
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX13139	..	443.00	742.46	5.2e-33
SIDS5/gcgdata/geneseq-genemb1/NA2001.DAT.AAS46237	..	438.50	732.67	1.8e-32
SIDS5/gcgdata/geneseq-genemb1/NA2002.DAT.ABA30341	..	438.50	670.32	5.5e-29
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20251	..	428.50	698.33	1.5e-30
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AA561567	..	425.50	733.40	1.7e-32
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AA561568	..	421.00	725.72	4.5e-32
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AA201425	..	420.50	650.01	7.4e-28
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20042	..	417.50	718.30	1.2e-31
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX13301	..	417.50	698.44	1.5e-30
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX13018	..	414.50	713.30	2.2e-31
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20030	..	404.50	669.36	6.2e-29
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX13092	..	403.50	693.49	2.8e-30
SIDS5/gcgdata/geneseq-genemb1/NA2001B.DAT.AAS2765	..	394.00	673.95	3.4e-29
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20031	..	393.50	676.12	2.5e-29
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX61686	..	390.50	670.88	5.1e-29
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20043	..	390.00	669.61	6.0e-29
SIDS5/gcgdata/geneseq-genemb1/NA1994.DAT.AAX55320	..	387.00	662.00	1.6e-28
SIDS5/gcgdata/geneseq-genemb1/NA1996.DAT.AAT42063	..	363.00	540.48	9.3e-22
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20040	..	368.50	612.58	9.0e-26
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX13040	..	358.50	598.98	5.1e-25
SIDS5/gcgdata/geneseq-genemb1/NA2001A.DAT.AAH54706	..	354.00	595.64	7.9e-25
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20056	..	351.00	602.61	3.2e-25
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX13168	..	351.00	592.53	1.2e-24
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20621	..	351.00	586.27	4.2e-24
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20041	..	349.00	596.27	7.3e-25
SIDS5/gcgdata/geneseq-genemb1/NA1998.DAT.AAX52194	..	337.00	551.65	2.2e-22
SIDS5/gcgdata/geneseq-genemb1/NA1999.DAT.AAX20057	..	329.50	565.31	3.9e-23

US-09-824-567-2 x AAD20238 .,

```
alignment_block:
  rvc = 00 004 400 0
```

Align seg 1/1 to: AAD20238 from: 1 to: 1799

```

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
101 ATGCGCAAGATATCAGTGGGAATCTGTATCACCATTTCTCTTAGCCTC 150
17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34
151 CGTAGCTCCAGGCTCGAAGAGTCCAGTCACTCTCTACATCTCGG 200
34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
201 GAGAACTCTCTATTATAAGATGCAACCCCGTCTTTAGATCCAGA 250
51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGlu 67
251 CAGTGGGACTCTCTTCAGAAATCAGCCTGTGCAACATATCTATGAGG 300
67 yLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAla 84
301 ATTAGTCAAGAAATAATCTTTCAGAAATATAGAGCTCTCTTCGAG 350
84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
351 AGACTACTCTCTTCTCGAGCGGACTCAGTTATCTTTAAACTGAA 400
101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117
401 TCAGCTTTTGGAGTAATGCGAGCCCTTAACAGCTGAGACTTATAGA 450
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPhe 134
451 ATCTTGGAAACAGTAGTACTCAAGAGTCTCAGGAATCTATGCTTTG 500
134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
501 CCTTGAATCCANNTAAATGTACAAAGATCCAGAGGGACACCTCTCC 550
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
551 ATAGACCAATTTGGAGTGACCTCTCTTAATGAATCTACACTTGTGTAC 600
167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184
601 CTGGAAATCCCAACCTCCATTTCTTAAACTTTTAGCTCTTCCAGTCT 650
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
651 TTTTCCCGCTTCATAAATCTCAAGAACCCCTGCATCCAAATCCAAATCTACCT 700
201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrp 217
701 ATAGCAAGCGAGCTTTCTATCTCTAAAAATATCAACAAAAACAATGAT 750
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThr 234
751 AAAACTCTCAAAAACCCCTCACTACTATAATCAAAAGTCAGTGGAACTA 800
234 ystThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
801 AAAGATTACGATTCACTTCATTCCTCCGATGCAACACAGCAGCAAACTA 850
251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgIle 267
851 TTTAATCAGGAAACTCAATTTGGCAAGGACCTCTCTTGGGGAGACGAT 900
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
901 TCCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACCTTACACTCT 950
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300

```

```

951 TTGATGTCCGAGGAACCTCATGCTCACTTCAATATCAATAAATTCCTCC 1000
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317
1001 CTCACAAATATGAGCTTTAGAGAGCTTTAGCATCAGCTTAGATAAGA 1050
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
1051 AGCTTTGTCTCAACTATATTCTTAGGCCGTGCAAAACTGCGCATCATC 1100
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
1101 TCCTACCTACAAATATTCTAGCTATCCGGAACATCAAAACACAGAGATG 1150
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGlu 367
1151 GCACACCCAGCTTTACGCTTAAAAACTCTTTAAAGAGCTTTAGAAGA 1200
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePhePro 384
1201 ACTCAAATCACTGTCTAAGATCTCGAATCTTAACTTATCTTTCTCC 1250
384 alSerSerSerAlaSerSerLeuValGlnLeuIleArgGlnTrp 400
1251 TTTCTCGTCAGCAAGTTCTTTACTAGTCCAACTTATACGAGACAGTGG 1300
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
1301 AAAGAAAGTTTAGGCTTCGCTATCCCTATTGTGCGAAAGAAATTTGCT 1350
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
1351 TCTCCAAAGCAGACCTATCTTCAGGAACTCTCTTAGCTACAGAGAGAT 1400
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
1401 GGTTCGACACTTTGCTGATCCTATGGCATTTCTAACGATCTTTGCTTAT 1450
451 ProSerGlyValProTyrAlaIleAsnHisLysAspPheLeuGluIle 467
1451 CCATCAGGAGTTCTCTTATGCAATCAACCAATAGGACTTCCTAGAAAT 1500
467 eLeuGlnAsnIleGluGlnGlnAspHisGlnLysArgSerGluLeuV 484
1501 TCTACAAACATAGAACAGCAAGATCACCACCAAGCTCGCAATTAG 1550
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
1551 TGTCCGACAGCTTCTTTACTAGACACCTTTTCATATTATTGAGCGCATC 1600
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517
1601 TACCACGAGCATTTCAATTTGCTATGAATAAAAACTTTCTAATCTAGG 1650
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
1651 AGTCTCAACCAACAGGAGTTGTGGACTTCGTTATGCTTAAGGAAAAAT 1696

```

seq_name: /SDS5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.AAX91990

seq_documentation_block:

ID AAX91990 standard; DNA; 1230025 BP.

XX AAX91990;

AC AAX91990;

DT 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; ss.

OS Chlamydia pneumoniae.
 PN WO9927105-A2.
 PD 03-JUN-1999.
 PF 20-NOV-1998; 98WO-IB01890.
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 PA (GEST) GENSET.
 XX Griffais R;
 PI WPI; 1999-357842/30.
 DR Genome sequence of Chlamydia pneumoniae
 PT Claim 1; Page 291-611; 1912pp; English.
 PS The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

alignment_scores:
 Quality: 2726.00 Length: 533
 Ratio: 5.124 Gaps: 1
 Percent Similarity: 99.812 Percent Identity: 99.812

alignment_block:
 US-09-824-567-2 x AAX91990 ..

Align seg 1/1 to: AAX91990 from: 1 to: 1230025

1 MetArgLysIleSerValGlyCysIleThrIleLeuLeuSerLeuSe 17
 |||||
 244480 ATGCGCAAGATATCAGTGGGAATCTGTATCACCATTCTCTTAGCCCTC 244529
 |||||

17 rValValLeuGlnGlyCysLysGluSerHisSerHisSerArgG 34
 |||||
 244530 CGTAGTCCTCCCAAGCTGCAAGAGTCCAGTCACTCTCTACATCTCGG 244579
 |||||

34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspArg 50
 |||||
 244580 GAGAACTCGCTATTATTAAGATGAAGATGACCCGCTTTAGATCCAA 244629
 |||||

51 GlnValArgLeuSerGluIleSerLeuValLysHisIleTyrGluG 67
 |||||
 244630 CAAGTCGACTTCTTTCAGAAATCAGCCCTGTGCAACATATCTATGAGG 244679
 |||||

67 yLeuValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
 |||||
 244680 ATTAGTTCAGAAATAATCTTTTCAGAAATATAGAGCTGCTCTTCAG 244729
 |||||

84 luAspTyrSerLeuSerAspGlyLeuThrTyrThrPheLysLeuLys 100
 |||||
 244730 AAGACTACTCTCTTCTCGGCGGACTCATTATATCTTTAAACTCAA 244779
 |||||

101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG 117
 |||||
 244780 TCAGCTTTTGGAGTATGCGGCCCTTAAACAGCTGAAGACTTTATAGA 244829
 |||||

117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
 |||||
 244830 ATCTTGGAAACAAGTAGCTACTCAAGAAGTCTCAGAAATCTATGCTTTG 244879
 |||||

134 laLeuAsnProIleLysAsnValArgLysIleGlnGlyHisLeuSer 150
 |||||
 244880 CCTTGAATCCAATTAATAATGTCAGAAAGATCCAAAGAGGACACTCTCC 244929
 |||||

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
 |||||
 244930 ATAGACCATTTTGGAGTCCACTCTCCTAATGAATCTACACTTGTGTAC 244979
 |||||

167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184
 |||||
 244980 CCTGGAATCCCAACCTCGCATTTCTTAAAACTTTTACCTCTCCAGTCT 245029
 |||||

184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
 |||||
 245030 TTTTCCCGTTCAATAATCTCAAGAACCCTGCAATCCAAATCTCTACCT 245079
 |||||

201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpI 217
 |||||
 245080 ATAGCAACGGAGCTTCTATCTCTAATAATATCAAAACAACAATGGAT 245129
 |||||

217 eLysLeuSerLysAsnProHisTyrTrpAsnGlnSerGlnValGluThrL 234
 |||||
 245130 AAACCTCTCAAAAAACCTCACTACTATATAATCAAGTCAGTGGAACTA 245179
 |||||

234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
 |||||
 245180 AAACGATTACGANTTCACTTCTTCGCGATGCAACACAGCAGCAAACTA 245229
 |||||

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgI 267
 |||||
 245230 TTTAATCAGGGAAAACTCAATTGGCAAGGACCTCTTGGGGAGAACCAT 245279
 |||||

267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
 |||||
 245280 TCCTCAAGAAACCCCTATCCAATTTACAGTCTAAGGGGCACTTACACTCT 245329
 |||||

284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
 |||||
 245330 TTGATGTGCGAGGAACCTCATGGCTCACCTTCAATATCAATAAATTCGCC 245379
 |||||

301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317
 |||||
 245380 CTCACCAATATGAAGCTTAGAAGCCCTTAGCATCAGCCTTAGATAAGGA 245429
 |||||

317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
 |||||
 245430 AGCTCTTGTCTCAACTATATTTCTTAGGCCGTGCAAAACCTGCCGATCATC 245479
 |||||

334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
 |||||
 245480 TCCTTACCTACAATATTCATAGCTATCCCGNACATCAAAAACAAGAGATG 245529
 |||||

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG 367
 |||||
 245530 GCACAGCGCCAGCTTACGCTAAAAAACTCTTTAAAGAAAGCTTTAGAA 245579
 |||||

367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
 |||||
 245580 ACTCCAATTCACCGCTAAGATCTCGAACACTCTTAATCTTATCTTTCCTCCG 245629
 |||||

384 alSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrp 400
 |||||
 245630 TTTCCTCGTCAGCAAGTCTTCTTACTAGTCCAACTTATACGAGACAGTGG 245679
 |||||

401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
 |||||
 245680 AAAGAAAGTTTGGGTTGCTATCCCTATTGTCGGAAGGAATTTGCTCT 245729
 |||||

417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434
 |||||


```
228 lnSerGlnValGluThrLysThrIleHisPheIleProAspAla 244
      :::::||||| ::::: ||| |||||:::|||||
251445 AAGACGAGTACTGTACAGAGATCTGTATACATCATCTTCATCAA 251396
      :::::||||| ::::: ||| |||||:::|||||
245 AsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyPr 261
      :::::||||| ::::: ||| |||||:::|||||
251395 CAAACTGCTTCTGCTTTATTCACCAAGGAAGCTAGATTGGCAAGTCT 251346
      :::::||||| ::::: ||| |||||:::|||||
261 oProTrpGlyGluArgLysProGlnGluThrLeuSerAsnLeuGlnSerL 278
      :::::||||| ::::: ||| |||||:::|||||
251345 CCCTGGGACATAGTATTCCGCAAGAACTTTAGCCACCAACAACAAAC 251296
      :::::||||| ::::: ||| |||||:::|||||
278 ysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPhe 294
      :::::||||| ::::: ||| |||||:::|||||
251295 GAGGAGCTCCCAATCATTTGATATATCTGCTACTTCTGCTTACATTT 251246
      :::::||||| ::::: ||| |||||:::|||||
295 AsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGluAlaLeuAl 311
      ||| ||| :::::||||| ::::: ||| |||||:::|||||
251245 AACACTGCCAAAGCCCTTTAGTCTATTCCTCAAGCTTCGCCAAGCTTGAG 251196
      :::::||||| ::::: ||| |||||:::|||||
311 aSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArg. 327
      : :::::||||| ::::: ||| |||||:::|||||
251195 TCTAGTTTAAACAAAGACCTCTGCCCTTGGCTTTGTTTNAACCT 251146
      ||||| :::::||||| ::::: ||| |||||:::|||||
328 AlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSerTrpProG 344
      ||||| :::::||||| ::::: ||| |||||:::|||||
251145 GCAGCTCTCTTATAGCAAGAGCCATCATCTTACCTTAACTTTAC 251105
      ||||| :::::||||| ::::: ||| |||||:::|||||
344 whisGlnGlnGluMetAlaGlnArgGlnAlaTyAlaLysLysLeuP 361
      ||||| :::::||||| ::::: ||| |||||:::|||||
251104 GCAGCTCTCTTATAGCAAGAGCCATCATCTTACCTTAACTTTAC 251055
      ||||| :::::||||| ::::: ||| |||||:::|||||
361 heLysGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGluHis 377
      :::::||||| ::::: ||| |||||:::|||||
251054 TAGAAGAAGCTCTGACTGAGCTTACATGACTATTAAGGATCTAGAGAA 251005
      :::::||||| ::::: ||| |||||:::|||||
378 LeuAsnLeuIlePheProValSerSerSerAlaSerSerLeuValG 394
      ||| ||| :::::||||| ::::: ||| |||||:::|||||
251004 TATCTCTTACCTTTTCCGCAAGCTCTACTGTGAACCTCACAGATGCTCA 250955
      ||||| :::::||||| ::::: ||| |||||:::|||||
394 nLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAlaIleProIleV 411
      ||||| :::::||||| ::::: ||| |||||:::|||||
250954 GATGTGCGGATCATGTGCGGAGAGTTTAGGAATTACTTCCCTTACT 250905
      ||||| :::::||||| ::::: ||| |||||:::|||||
411 alGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsn 427
      ||||| :::::||||| ::::: ||| |||||:::|||||
250904 GCGGAAAGAATATGCTTTGTTGAAACGATCTAATAGCAATCTTTC 250855
      ||||| :::::||||| ::::: ||| |||||:::|||||
428 SerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAla 444
      :::::||||| ::::: ||| |||||:::|||||
250854 TTTATGCTATAGTGGTGGTGTTCGACGCTTCTGACGCTTTAGCGTT 250805
      ||||| :::::||||| ::::: ||| |||||:::|||||
444 eLeuThrIlePheAlaTyProSerGlyValProProTyAlaIleAsnH 461
      ||||| :::::||||| ::::: ||| |||||:::|||||
250804 TCTTTCATTTTCTCC...TCGAAGAGATCAACCTTTATGCTTTACAG 250758
      ||||| :::::||||| ::::: ||| |||||:::|||||
461 isLysAspPheLeuGluLeuLeuGlnAsnIleGluGlnGlnAspHis 477
      :::::||||| ::::: ||| |||||:::|||||
250757 ATCTCTAATTTGATCACTGATCTCTCTATAGAACGGAAAAAACCT 250708
      ||||| :::::||||| ::::: ||| |||||:::|||||
478 GlnLysArgSerGluLeuValSerGlnAlaSerLeuTyLeuGluThrPh 494
      ||||| :::::||||| ::::: ||| |||||:::|||||
250707 CAAAAGCTCAGCTTAAATTCGAGAGCTTCTCTATACATAGAAAGACA 250658
      ||||| :::::||||| ::::: ||| |||||:::|||||
494 eHisIleIleGluProIleTyHisAspAlaPheGlnPheAlaMetAsn 511
      :::::||||| ::::: ||| |||||:::|||||
250657 AAAGCATAGAACCCCTCTTACGACGCTGTTCCATTATACAAATA 250608
      ||||| :::::||||| ::::: ||| |||||:::|||||
511 ysLysLeuSerAsnLeuGlyValSerProThrGlyValValAspPheArg 527
      :::::||||| ::::: ||| |||||:::|||||
250607 ATAACTTCTTTTGTAGACTACATCTTCAGGCTTCTGATATGCGG 250558
      ||||| :::::||||| ::::: ||| |||||:::|||||
```

```
528 TyrAlaLysGluAsn 532
      |||||:::|||||
250557 TATGCTAAAAACTCT 250543
```

seq_name: /SDS5/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA518750

seq_documentation_block:

ID AAS18750 standard; DNA; 1787 BP.

XX AAS18750;

XX DT 26-MAR-2002 (first entry)

XX Chlamydomonada pneumoniae gene encoding an ATP-binding cassette.

XX ATP binding cassette; secretory locus open reading frame; endopeptidase;
XX secretory locus ORF; protease; metalloprotease; Clp protease Atpase;
XX Clp protease subunit; transglycolase/transpeptidase; ClpC protease;
XX thioesterase; Chlamydia infection; antibacterial; ds.

XX Chlamydomonada pneumoniae CWL029.

XX WO200185972-A2.

XX PD 15-NOV-2001.

XX PF 08-MAY-2001; 2001WO-CA00653.

XX PR 30-MAY-2000; 2000US-202672P.

XX PR 16-JUN-2000; 2000US-207852P.

XX PR 16-JUN-2000; 2000US-211796P.

XX PR 16-JUN-2000; 2000US-211797P.

XX PR 16-JUN-2000; 2000US-211798P.

XX PR 16-JUN-2000; 2000US-211801P.

XX PR 16-JUN-2000; 2000US-212044P.

XX PR 26-SEP-2000; 2000US-235335P.

XX PR 26-SEP-2000; 2000US-235361P.

XX PR 26-SEP-2000; 2000US-235398P.

XX (AVET) AVENTIS PASTEUR LTD.

XX Murdin AD, Oomen RP, Wang J, Dunn P;

XX WPI; 2002-049447/06.

XX P-PSDB; AAU09430.

XX Vaccine useful for immunising mammals against chlamydia infections,
XX comprises vectors having sequences of ATP binding cassette gene,
XX secretory locus open reading frame gene of chlamydia -
XX Claim 1; Fig 1; 355pp; English.

XX The present invention relates to the isolation of Chlamydomonada
XX pneumoniae strain CWL029 genes and their encoded proteins. The genes of
XX the invention encode an ATP binding cassette gene, a secretory locus
XX open reading frame (ORF), an endopeptidase, a protease, a
XX metalloprotease, Clp protease Atpase, a Clp protease subunit, a
XX transglycolase/transpeptidase, a ClpC protease, or thioesterase. The
XX genes of the invention can be used in a vector as a vaccine for the
XX prevention and treatment of Chlamydia infections. Also described are
XX B- and T-cell epitopes from the proteins of the invention which can be
XX used as Chlamydia antigens. AAS18750-AAS18759 represent the C. pneumoniae
XX genes of the invention.

SQ Sequence 1787 BP; 628 A; 436 C; 261 G; 462 T; 0 other;

alignment_scores:

Quality: 755.50 Length: 522

Ratio: 2.116 Gaps: 9

Percent Similarity: 68.391 Percent Identity: 33.908

alignment_block:

```

seq_name: /SID85/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA51864
seq_documentation_block:
ID   AAA51864 standard; DNA; 6300 BP.
XX
XX
AC   AAA51864;
XX
XX
DT   31-OCT-2000 (first entry)
XX
XX
B. subtilis opp operon.
XX
XX
KW   Opp operon; SpoOK; oligopeptide permease; sporulation; ABC trans-
KW   ATP-binding cassette transporter; mutation; protein production
KW   oppB; oppC; oppD; oppF; ds.

```


1339 AAGACGGTTCTTTACATGTTGACCCGATTCAGGAGTGTATGGTACAAA 1388
 294 PheAsnIleAsnLysPheProLeuAsnMetLysLeuArgGluAlaLe 310
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
 1389 TTCAACACTGAAGCTAAGCCATAGACAACGTCAATATCGTAAAGCTTT 1438
 310 uAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyA 327
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1439 AACATATCGCTTGACCGTCAGTCAGTCAGTGTAAACAGCTTACGCAAGG 1488
 327 rGAlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrPro 343
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1489 AGCAAAATCCCGCAATAGCTGCGAGTCCGCTACATGAAGGATTTGAG 1538
 344 GluHisGlnLysGlnGluMetAlaGlnArgGlnAla...TyrAlaLysLY 359
 :
 1539 GATACCAAGAGGATGACTTCAAGAGCAATGATGTCAAAACACAGCAAGA 1588
 359 sLeuPheLysGluAlaLeuGluGluLeuGlnIleThr...AlaLysAspL 375
 :
 1589 ATACCTTGAAAAGAGCCCTAAGAAATGGCGCTTAAGCAAGCATCTGAT 1638
 375 euGluHisLeuAsnLeuIlePheProValSerSerAlaSerSerLeu 391
 || : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1639 TGCAAAATCAATGATGCTTACAC...ACTGATGACGACACACGCGAA 1685
 392 LeuValGlnIleArgGluThrLysLysGluSerLeuGlyPheAlaI 408
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1686 ATGCTCAGCAGTACAAAGAAATGTGAAGAAATTTAGCGGTGTGT 1735
 408 eProLysValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerG 425
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1736 TGAGCTTGATACTCAGAGTGAATGCTATATGATGAAGTCCACAGCC 1785
 425 LysPheSerLeuAlaThrGlyClyTrpPheAlaAspPheAlaAspPro 441
 : : : : : : : : : : : ||||| : : : : : ||||| : : : : :
 1786 AGATTTCAATCGCGGTATGGTGGCTGGCGACTTCAATGATCCT 1835
 442 MetAlaPheLeuThrIlePheAlaTyrProSer...GlyValProTyr 457
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1836 ATCACTTCCTTGATGTTCCCGCAAAACAGGAGGAATAACGATAC 1885
 457 rAlaIleAsnHisLysAspPheLeuGluIleLeuGlnAsnIleGluGln 474
 : : : : : : : : : : : ||||| : : : : : ||||| : : : : :
 1886 AGCTGGGAAATCCAGAAATTCAAAAGCTTCTGAATCAGTCACAACTG 1935
 474 LuGlnAspHisGlnLysArgSerGluLeuValSerGlnAla...SerLeu 489
 || ||||| : : : : : ||||| : : : : : ||||| : : : : :
 1936 AACAGATAAAACAAACGTCGAGCTGCTGAAAAAGCAGAAGGTATT 1985
 490 TyrLeuGluThrPheHisIleLeuGluProIleTyrHisAspAlaPheG 506
 : : : : : : : : : : : ||||| : : : : : ||||| : : : : :
 1986 TTCATTGATGAATGCG...GTGCCCATCTATTTCATCTACTGATAC 2032
 506 nPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProThrGlyV 523
 : : : : : : : : : : : ||||| : : : : : ||||| : : : : :
 2033 TTGGTACAGATGAACAACTTAAAGGTGTTATCATCGCAGGTACTGGTG 2082
 523 alValAspPheArgTyrAla 529
 ||||| |||||
 2083 AGTTTATTTCAGAACGCA 2102

seq_name: /SIDS5/gcgdata/geneseq/emb1/NA1999.DAT.AAX61764

seq_documentation_block:

ID AAX61764 standard; DNA; 1536 BP.

AC AAX61764;

DT 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein coding sequence, t606.nt.

XX

KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
 XX Borrelia burgdorferi.
 OS
 XX WO9859071-A1.
 XX
 PD 30-DEC-1998.
 XX
 XX 18-JUN-1998; 98WO-US12718.
 XX
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 XX
 XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX WPI: 1999-189980/16.
 DR P-PSDB; AAY20067.
 DR
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases
 PT caused by Borrelia, particularly Lyme disease
 XX
 PS Claim 1; Page 182-183; 275pp; English.
 XX
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.
 XX
 SQ Sequence 1536 BP; 631 A; 246 C; 256 G; 403 T; 0 other;

alignment_scores:

Quality: 505.50 Length: 485
 Ratio: 1.620 Gaps: 15
 Percent Similarity: 64.330 Percent Identity: 28.041

alignment_block:

US-09-824-567-2 x AAX61764 ..

Align seg 1/1 to: AAX61764 from: 1 to: 1536

38 IleAsnIleArgAspGluProArgSerLeuAspProArgGlnValArgLe 54
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
 43 ATAAGCTTGGGAGCAGACCAAGCAGTCTTGACCTCAATTAGCAGAGGA 92
 54 uLeuSerGluIleSerLeuValLysHisIleTyrGluGlyLeuValGlnG 71
 :
 93 TAATGTGCGCATCAAAAATGATGTACACAAATGTTAGAGGGATTTACAG 142
 71 LuAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAspTyrSer 87
 :
 143 GAGATCCTTAATACAGGGGAAATAAACCGGACCTTGCAAAAGGGTGGAT 192
 88 LeuSerSerAspGlyLeuThrTyrThrPheLysLeuLysSerAlaPhe.. 103
 :
 193 ATTTCTTCTGATGAACAGCTTACACTTTAACCTTAAGAGAAAAATCAC 242
 104 TrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerTrpL 120
 ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
 243 TTGGAGTACGGGAGTTCATCACTACGACAGGAATTAGAAAAATCTTATC 292
 120 ySlnValAlaThrGlnGluValSerGlyIleTyrAla...PheAlaLeu 135
 :
 293 TTAGAAATTTAAATAAGAACTGGCTCAAACTACGTTGAATGGTTAAA 342

421 AspLeuSerSergIysAsnPhseSerLeuAlaThrGlyGlyTrpPheAlaAs 437
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1183 ACTAAGGCCAAATGGGAATTATGAATAAGCAAGACGAGGTGATAGCGA 1232

437 pPheAlaSprProMetAlaPheLeuThrIlePheAlaTyrProSerGlyV 454
|||:::|||||:::|||||:::|||||:::|||||:::|||||
1233 TTATGCTGATCCTTTGACATTTTAAGCATATTACA.....CAAGSAT 1276

454 alProProTyraLala.....IleAsnHisLyAspPheLeuGlulle 467
:::|||||:::|||||:::|||||:::|||||:::|||||
1277 ACACACAAFTCTCTCATTAATTACTCAAACCCAGAATACACGAACCT 1326

468 LeuGlnAnlleGluInGluInAspHisGlnLysArgSerGluLeuVa 484
:::|||||:::|||||:::|||||:::|||||:::|||||
1327 ATAAAGAAAATCCGACCTTGAGCTTGATCCAATAAAAGACAAAGACATT 1376

484 lserGlnAla...SerLeuTyrLeuGluThrPheHisIleileGluProI 500
: |||||:::|||||:::|||||:::|||||:::|||||
1377 AAGACAGCAGAGAGAGATAATTATGAAAAAGATTTCCTCAATAGCACCAA 1426

500 leTyr 501
|||||
1427 TATAC 1431

seq_name: /SIDS5/gcgdata/geneseq/geneseqn_emb1/NA1999.DAT:AAx6176
seq_documentation_block:
ID AAx61763 standard; DNA; 1587 BP.
XX
AC AAx61763;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, f606.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection
OS Borrelia burgdorferi.
XX
XX WO9859071-A1.
XX
PN 30-DEC-1998.
PD
XX
PF 18-JUN-1998; 98WO-US12718.
XX
XX 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
XX {HUMA-} HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
PI WPI; 1999-189980/16.
DR P-PSDB; AAY20066.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop products for the diagnosis, prevention and treatment of diseases caused by Borrelia, particularly Lyme disease
PT
PT Claim 1; Page 182; 275pp; English.
PS
PS This sequence encodes a Borrelia burgdorferi (Bb) protein of invention, which is suitable for use in a vaccine. The Bb polypeptide can be used in vaccines for eliciting protective antibodies against the Borrelia genus, particularly for the use against Lyme disease in humans and animals. They can be used for preventing or attenuating infection caused by a member of the Borrelia genus. The product can be used for detection of members of the Borrelia genus.
CC
XX Sequence 1587 BP; 648 A; 251 C; 259 G; 429 T; 0 other.

```

alignment_scores:
  Quality: 505.50      Length: 485
  Ratio: 1.620         Gaps: 15
  Percent Similarity: 64.330  Percent Identity: 28.041

alignment_block:
US-09-824-567-2 x AAX61763
Align seg 1/1 to: AAX61763 from: 1 to: 1587

38  lleAsnIleArgAspGluProArgSerLeuAspProArgGlnValArgIle 54
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
94  ATAAGCTTGGAGCAGACGCAAGCAGCTCTGACCTCAATAGCAGAGA 143
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
54  uLeuSerGluIleSerLeuValLysHisIleTyrGluGlyLeuValGlnG 71
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
144  TAATGTCGCATCAAAATGATTGACACAATGTTAGAGGATGTTTACAG 193
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
71  luAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAspTyrSer 87
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
194  GAGATCTTAATACAGGGGAATTAACCGGACTTGCRAAAGGGTGGAT 243
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
88  LeuSerSerAspGlyLeuThrTyrThrPheLysLeuLysSerAlaphe... 103
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
244  ATTCTTCTGTGAGAACAGTTTACACATTTAACTAAGAGAAAAATCAC 293
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
104  .TrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerTrpL 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
294  TTGGAGTACGAGCTGCATCACTGAGAGGAATTTAGAAAATCTTATC 343
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
120  ySglnValAlaThrGlnGluValSerGlyIleTyrAla...PheAlaLeu 135
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
344  TTAGAATTTTAATAAGAAACTGGCTCAAGTAGCTTGAATGGTAAA 393
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
136  AspProIleLysAsnValArgLysIleGlnGluGlyHisLeuSerIleAs 152
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
394  TCGTAATTAATAATGTCAAAATATTTTGTGACCAAGTGAAGTACTG 443
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
152  pHisPheGlyValHisSerProAsnGluSerThrLeuValValThrLeuG 169
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
444  TGAAGTTGGAATTAGACGATTCATGAAAAACATTAGAAATAACACTGG 493
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
169  luSerProThrSerHisPheLeuLysLeuAla...Leu 181
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
494  AATCAACAAAACCTTATTTTATGATATGTTAGTACCAATCATTTAT 543
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
182  ProValPhePheProValHisLysSerGlnArgThrLeuGlnSerLysSe 198
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
544  CCAGTA.....CCAGTTCATGTTACCGAAAAGTATGCACAAACTGGAC 587
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
198  rLeuPro.....IleAlaSerGlyAlaPheTyrProLysAsnIleL 212
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
588  AAGCCCGCAACATGTCGACAGTGGCTCTTTTAAATTAAGAAAGAA 637
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
212  ysGlnLysGlnThrIleLysLeuSerLysAsnProHisTyrAsnGln 228
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
638  TTCCTAACGAAAAATATGCTTTTGAATAAATAACAAATACTACGACTCA 687
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
229  SerGlnValGluThrLysThrIleThrIleHisPheIleProAspAlaAs 245
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
688  AATGAGTAGAATTAAGAGAGATTTACATTTTACACAAATAGCAGCTC 737
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
245  nThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyProp 262
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
738  AACAGCTATAAATGATGAAATGAAGAGCTAGAT.....GCAA 778
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
262  roTrpGlyGluArgIleProGlnGlnThrLeuSerAsnLeuGlnSerLys 278
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
779  TTTTGGTTC...ATCCCCCAGATCAATCAAAATCTAAATTAAGA 825
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
279  GlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPheAs 295
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: AAX20248

seq_documentation_block:

ID AAX20248 standard; DNA; 910715 BP.

XX AAX20248;

XX 04-MAY-1999 (first entry)

XX Borrelia burgdorferi polynucleotide sequence #1.

XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 XX epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.

421 AspLeuSerSerGlyAsnPhSerLeuAlaThrGlyGlyTrpPheAlaAs 437
 572850 ACTAAGCAAAATGGAATTTAGAAATAGCAAGAGGATGATGAGCGGA 572801
 437 pPheAlaAspProMetAlaPheLeuThrIlePheAlaTyrProSerGlyV 454
 572800 TTATGCTGATCCTTGATGATTTTAAAGCATATTCACA.....CAAGGAT 572757
 454 alProProTyrAla.....IleAsnHisLysAspPheLeuGluIle 467
 572756 ACACACAATTCATCTCATATTAATCAACCCAGATATCAACGAACCT 572707
 468 LeuGlnAsnIleGluGlnGluGlnAspHisGlnLysArgSerGluLeuVa 484
 572706 ATAAGAAATCCGACCTTGAGCTTGCATCCATAAATAAGACAGACATTT 572657
 484 lSerGlnAla...SerLeuTyrLeuGluThrPheHisIleIleGluProI 500
 572656 AAGACACAGACAGAGATATTAATTTGAAAAAGATTTTCCAAATAGCACCA 572607
 500 lTyr 501
 572606 TATAC 572602

seq_name: /STD5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAAX20020

seq_documentation_block:

ID AAX20020 standard; DNA; 1659 BP.

XX AAX20020;

XX AC AAX20020;

XX DT 20-APR-1999 (first entry)

XX DE Enterococcus faecalis gene EF012.

XX KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

XX OS detection; attenuation; antigenic; ss.

XX PN W09850554-A2.

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98MO-US08959.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;

XX PT WPI: 1999-070095/06.

XX DR P-PSDB: AAU00030.

XX PS New isolated Enterococcus faecalis polynucleotides - used to develop

XX PT products for the detection of Enterococcus and for use in vaccines

XX PT for prevention or attenuation of Enterococcus infection

XX PS Claim 1; Page 89-90; 301pp; English.

XX CC The present sequence represents a gene isolated from

XX CC Enterococcus faecalis. The present invention describes genes, proteins

XX CC and antigenic polypeptides isolated from E. faecalis. The proteins can

XX CC be used in vaccines for preventing or attenuating an infection caused

XX CC by a member of the Enterococcus genus in an animal. They can also be

XX CC used for detecting Enterococcus antibodies in a sample. The nucleotide

XX CC sequences can be used for detecting Enterococcus nucleic acids.

XX CC Products from the present invention can also be used for screening

XX CC compounds to identify agonists and antagonists of E. faecalis protein

CC activity.
 XX
 SQ Sequence 1659 BP; 582 A; 305 C; 359 G; 413 T; 0 other;
 alignment_scores:
 Quality: 455.00 Length: 539
 Ratio: 1.346 Gaps: 15
 Percent Similarity: 62.709 Percent Identity: 25.232
 alignment_block:
 US-09-824-567-2 x AAX20020 ..

Align seg 1/1 to: AAX20020 from: 1 to: 1659

4 IleSerValGlyIleCysIleThrIleLeuLeuSerLeuSerValValle 20
 16 ATGAATTTGGGGAAAAAGTAGTAGTTTATTGATTCACACAGGGTTTCTTT 65
 20 uGlnGlyCys.....LysGluSerSerHis 29
 66 AGCCGATGTGGCGGAACCAAGAGCGGCGAGAAAGTAGATTTCGGAA 115
 29 erSerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArg 45
 116 ATTAGCAGCTGAACAAAAATCAGTATTAGTTTACCTGCACCAATCTCA 165
 46 SerLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuVally 62
 166 ACATTTGGATACACACAAACACAGATAAAATACCTTTACATGCGACA 215
 62 shisileTyrGluGlyLeuValGlnGluAsnAsnLeuSerGlyAsnIleG 79
 216 ACATTTATTTGAAGGCTTTATCGGTTTGTATGATGATGATGATGATG 264
 79 luProAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyr 95
 265 ..CCAGCTCTAGCTAAAGATGTCAGATAGTACGATGGCGGCAAGTAC 312
 96 ThrPheLysLeuLysSerAlaPhe...TrpSerAsnGlyAspProLeuTh 111
 313 CACTTTACCTTGGGAGGGGATTAAAGTGGAGCAAGCGGCGCAATCAC 362
 111 rAlaGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValS 128
 363 GGCCCAAGATTTTGTATTCTTGGAAAAAAGCTGGTGACACAGCGGCGA 412
 128 erGlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysile 144
 413 TTGGACCGAATGCTTATTACTAGACAGTGTAAAAATAGTTTGAATA 462
 145 GlnGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGl 161
 463 CGCAACGGTGAAGAGTCAAGTCAGTGAATAGGATTTTCAGCCCGCAATGA 512
 161 uSerThrLeuValValThrLeuGluSerProThrSerHisPheLeuLysL 178
 513 CAAGAATTTATTTGTGAATTAACACAGCGCCCAACCTTCTCTTACAG 562
 178 euLeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLeu 194
 563 TCGTTTCGATGCTTGGTTAGCGCCCAAAAAATCAAAAAATTTGTCGAGCG 612
 195 GlnSerLysSerLeuPro.....IleAlaSerGlyAl 205
 613 CAAGCAAGATTAAGCTTGGATAGTGAACATTTACTTTATAGCGGGCC 662
 205 aPheTyrProLysAsn.....IleLysGlnLysGlnTrpIleLysLeuS 220
 663 ATTTACGCTAGCAATTTGGATGCGCACTTACACTTGG...ACATGCA 709
 220 erLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrLysThrile 236

```

710 AAAAAATCAGAAATACTATGATCGGATCAAGTGAAGTGAAGAGTT 759
237 ThrileHisPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnG1 253
      ::::: ||| :::: |||||::: ::::: |||||:::
760 GCGGTTAGCACATCAAGACAGATTAATCTGGGATTAATCTATCAAGT 809
      ::::: ||| :::: |||||::: ::::: |||||:::
253 nGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgIleProGlnG 270
      ::::: ||| :::: |||||::: ::::: |||||:::
810 GAATGAACACTAGACTTA.....GTTCCGATTACGGAC 841
270 luThrLeuSerAsnLeuGln...SerLysGlyHisLeuHisSerPheAsp 285
      ::::: ||| :::: |||||::: ::::: |||||:::
842 AATATGTTCAAGATATCAAGATCAAGGCTATGTCAGTCAATCCAGAT 891
286 ValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLys...PheProLe 301
      ||||| :::: |||||::: ::::: |||||:::
892 GTGGCC...AATCTCTTCTAGATTCAACAAAAAGAGAACCCCAT 938
301 wasAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluA 318
      ||||| :::: |||||::: ::::: |||||:::
939 AGCGAATGTTCAATTTACGAAAGCGATTGGCCAAAGCAATGATAAGAG 988
318 laLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeu 334
      ||||| :::: |||||::: ::::: |||||:::
989 CCTTAACACAAGTGTCTTAACAGATGGGTCAAAACCCCTTAACGGATT 1038
335 LeuProThrAsnIleHisSerTrpProGluHisGlnLysGlnGluMetAl 351
1039 ATTCCAAGTAACTTTATGCGAATCCAGAACGGATGAAGATTTCGAGC 1088
351 aGlnArgGlnAlaTyrrAlaLysLysLeuPheLysGluAlaLeuGluL 368
1089 TTACAGTGGCAATATTGAAATGACGTCAAAAGAGCTCAAGCTGAAT 1138
368 euGlnIleThrAlaLysAspLeu...GluHisLeuAsnLeu...IlePhe 382
1139 GGACGAAAGCCCAAGCGGATGTCGTAAAAAAGTGAACCTTCATGCTG 1188
383 ProValSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluG1 399
1189 GCGGACAGACAGATCAAGGAAAGCAATGCTGAATATGTTCAAGTCA 1238
399 nTrpLysGluSerLeu...GlyPheAlaIleProIleValGlyLysGluP 415
1239 GTTCAGCAAAATTCGCGAGTTTGAATATACCATTTTCATCGCAACCA 1288
415 heAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThr 431
1289 GTAATAATGTGAACCAATCGCGACGTGAAAAAATTAAGTGTCTCTT 1338
432 GlyGlyTrpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePh 448
1339 TCAGATGGATTCGCGGAGTGTAGTACCTTACTTTAACTTATA 1388
448 eAlaTyrrProSerGlyValProProTyrrAlaIleAsnHisLys..... 462
1389 TGCAGGAGATCAAGTTAC...AATAGCGCAATTCATATATGCAAAAT 1435
463 .....AspPheLeuGluIleLeuGlnAsnIleGluGlnGluAspHis 477
1436 ACGACCAATTTGGTAGAGAGCGACGACGATTAATGCCAATATCCAGAG 1485
478 GlnLysArgSerGluLeuValSerGlnAlaSerLeuTyrrLeuGluThrPh 494
1486 AAACAGTTTCAGATACAAAGAGAGGAGACATCTTTGTTGAACCAAGA 1535
494 eHisIleIleGluProIleTyrrHisAspAlaPheGlnPheAlaMetAsnL 511
1536 TGCTGCCAAGTACCGCTGTATCAAGTGCCTCAATTTATCTAATCAATC 1585
511 ysLysLeuSerAsnLeu 516
1586 CTAATTTGAAGGCATT 1602

```

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.AAX13087

seq_documentation_block:

ID AAX13087 standard; DNA; 11739 BP.

XX AAX13087;

AC 19-MAR-1999 (first entry)

DT Enterococcus faecalis genome contig SEQ ID NO:150.

DE Enterococcus faecalis; contig: detection; Enterococcal infection;

XX Enterococcus faecalis; computer readable medium; ds.

KW Enterococcus faecalis.

OS W09850555-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-US08985.

PF 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

XX Claim 1; Page 862-868; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC AAX12938 to AAX13919 represent these nucleotide sequences which are

CC primary nucleotide sequences, also known as contigs. The computer-based

CC system can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

XX Sequence 11739 BP; 3917 A; 2161 C; 2537 G; 3118 T; 6 other;

alignment_scores:

Quality: 455.00 Length: 539

Ratio: 1.346 Gaps: 15

Percent Similarity: 62.709 Percent Identity: 25.232

alignment_block:

US-09-824-567-2 x AAX13087 ..

Align seg 1/1 to: AAX13087 from: 1 to: 11739

4 lleSerValGlyIleCysIleThrIleLeuLeuSerLeuSerValValle 20

2994 ATGAAATGGGAAAAAGTAGTAGTTGATTCACACAGGGTTCTTT 3043

20 uGlnGlyCys.....LysGluSerSerHis 29

[illegible]

XX 04-MAY-1998; 98WO-US08959.
 XX PF
 XX 14-NOV-1997; 97US-0066009.
 XX PR
 XX 06-MAY-1997; 97US-0044031.
 XX PR
 XX 16-MAY-1997; 97US-0046655.
 XX PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX Bailey C, Choi GH, Bromocky J A, Kunsch CA;
 XX PI
 XX WPI: 1999-070095/06.
 XX DR
 XX P-PSDB; AAY00031.
 XX DR
 XX PT
 XX New isolated Enterococcus faecalis polynucleotides - used to develop
 XX products for the detection of Enterococcus and for use in vaccines
 XX for prevention or attenuation of Enterococcus infection
 XX PT
 XX Claim 1; Page 90-91; 301pp; English.
 XX PS
 XX CC
 XX The present sequence encodes an antigenic polypeptide fragment
 CC isolated from Enterococcus faecalis. The present invention describes
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
 CC The proteins can be used in vaccines for preventing or attenuating an
 CC infection caused by a member of the Enterococcus genus in an animal.
 CC They can also be used for detecting Enterococcus in a sample.
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic
 CC acids. Products from the present invention can also be used for
 CC screening compounds to identify agonists and antagonists of E. faecalis
 CC protein activity.
 XX CC
 XX Sequence 1585 BP; 559 A; 296 C; 336 G; 394 T; 0 other;

alignment_scores:
 Quality: 452.50 Length: 519
 Ratio: 1.380 Gaps: 17
 Percent Similarity: 63.198 Percent Identity: 26.590
 alignment_block:
 US-09-824-567-2 x AAX20021 ..
 Align seg 1/1 to: AAX20021 from: 1 to: 1585
 22 GlyCysLysGluSerSerHisSerThrSerArgGlyGluLeuAla... 37
 8 GGAACCAAAAGAACCGCAGAGAAAGTAGATTGCG...GGAAATTACACC 54
 38ileAsnileArgAspGluPro.....ArgSerLeuAsp 49
 55 TGAACAAAATACGATTAGTTAGTACCTGCACCAATCTCAACATTGGATA 104
 49 roArgGlnValArgLeuSerGluileSerLeuValLysHisIleTy 65
 105 CAACACAAACACAGATAAAATACCTTTACATGCGCAACATTATTT 154
 66 GluGlyLeuValGlnGluAsnLeuSerGlyAsnileGluProAlaLe 82
 155 GAAGGCTTTATCGGTTTATGATGATGATGATGATGATGATGATGAT 201
 82 uAlaGluAspTySerLeuSerSerAspGlyLeuThrThrPheLysL 99
 202 AGCTAAGATGTCAGATTAGTACGATGATGATGATGATGATGATGAT 251
 99 eulysSerAlaPhe...TrpSerAsnGlyAspProLeuThrAlaGlu 114
 252 TGGGGGGGGGATTAAGTGGAGCAACGGCGCAATCAACGCCCAAGAT 301
 115 PheileGluSerTyPlyGlnValAlaThrGlnGluValSerGlyIleTy 131
 302 TTTGTTTATTTCTGGAAAAACCTGGTCACACACGACGATGGACCGAA 351
 131 rAlaPheAlaLeuAsnProLysAsnValArgLysIleGlnGlyH 148

352 TGCCTATTACTAGACAGTCTTAAATAATAGTTTGAATAGCAAGGTG 401
 148 isLeuSerileAspHisPheGlyValHisSerProAsnGluSerThrLeu 164
 402 AAAAGTCAGTCGATGAATTAGGATTTAGCCCGGATGACAAAGATTC 451
 165 ValValThrLeuGluSerProThrSerHisPheLeuLysLeuAlaLe 181
 452 ATTGTTGAATTAACACAGCCCAACCTCTTCTTACGATCGTTTCGAT 501
 181 uProValPhePheProValHisLysSerGlnArgThrLeuGlnSerLys 198
 502 TGCTTGTTAGCGCCCAAAATTTGTCGAAGCCCAAGGCAAG 551
 198 erLeuPro.....ileAlaSerGlyAlaPheTyPro 208
 552 ATTACGCTTGGATAGTGAACATTTACTTTATAGCGGCGCATTTACGCTA 601
 209 LysAsn.....ileLysGlnLysGlnTrpIleLysLeuSerLysAsnPr 223
 602 GCCAATTGGGATCGGACTCAGACTTGG...ACATGAAAAAAATCC 648
 223 oHisTyTyAsnGlnSerGlnValGluThrLysThrIleThrIleHisP 240
 649 AGAATCATGATGATGCGGATCAAGTGAAGTGAAGTGGGTTAGCA 698
 240 heileProAspAlaAsnThrAlaAlaLysLeuPheAsnGlnLysLeu 256
 699 CAATCAAGAACATTAATCTGGGATTAATCATCAAGTGAATGAACATA 748
 257 AsnTrpGlnGlyProTrpGlyGluArgIleProGlnLysLeuSe 273
 749 GACTTA.....GTTGCAATTACCGACATATGTTCA 780
 273 rAsnLeuGln...SerLysGlyHisLeuHisSerPheAspValAlaGlyT 289
 781 ACAATATCAAGATGATGATGATGATGATGATGATGATGATGATGAT 827
 289 hrSerTrpLeuThrPheAsnIleAsnLys...PheProLeuAsnMet 304
 828 ACTACTTCTTAGATTTCACAAAAGAGAGAGAGAGAGAGAGAGAGAG 877
 305 LysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSe 321
 878 CATTTACGAAAGAGAGATGCGCAAGCAATTAAGAGAGCTTAAACACA 927
 321 rThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeuLeuProThrA 338
 928 AGTGTCTTAACAGTGGGTCAAAACCCCTTAACGAGTGTGATTCAGTA 977
 338 snileHisSerTyProGluHisGlnLysGlnGluMetAlaGlnArgGln 354
 978 AACTTTATGCGAATCCAGAAACGATGATGATGATGATGATGATGATG 1027
 355 AlaTyAlaLysLysLeuPheLysGluAlaLeuGluGluLeuGlnIleTh 371
 1028 GAATATTGAAAAATGACGTCAAAAAGCTCAAGTCAATGGAGAGAGC 1077
 371 rAlaLysAspLeu...GluHisLeuAsnLeu...ilePheProValSerS 386
 1078 CCAAGCGGATGTCGGTAAAAAGTGAACCTTTCAATGCTGCGCGCAGACA 1127
 386 erSerAlaSerSerLeuValGlnLysGlnIleArgGluGlnTrpLysGlu 402
 1128 CAGATCAAGGAAACCAATTCGTAATATGTTCAAGTCAAGTTCAGTAAG 1177
 403 SerLeu...GlyPheAlaIleProIleValGlyLysGluPheAlaLeuLe 418
 1178 AATCTGCGAGGTTTAGAAATACCATTTTCATCGCAACCAAGTAATAATGT 1227
 418 uGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyTrp 435

```

alignment_scores:
  Quality: 451.50      Length: 501
  Ratio: 1.360        Gaps: 15
  Percent Similarity: 66.267      Percent Identity: 36.347

alignment_block:
  US-09-824-567-2 x AAX20253/rev ..

Align seg 1/1 to reverse of: AAX20253 from: 1 to: 26811

1 MetargLysileSerValcIlyleCysIleThrIleLeuLeuSerLeuSe 17
  :::::|||||:::|||||::: :::::|||||::: :::::
14530 ATAAAAGGTTAAAGGTGATTATTTCTCAATTAATTTACTATTTC 14461

17 rValValLeuIngIlyCysLysGluSerSerHisSerThrSerArgC 24

```

14480 TTGTGTT.....AATGAAAGCTAATAGAAACAAAT 14452

34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerIleuAsnProArg 50

[illegible]

14451 TGGTTTAAAGCTAAATATTGGAAGTGAGCCTGCTACTTTAGATGCTCAA 14402

51 GlnValArgLeuLeuSerGlu]eserLeuValIleuValCysMet

[illegible]

14401 TTAATAAACGATACGGTTGGATCAGGGATTGTAAGCCAAATGTTTCTTGG 14352

67 vLeijVa[G]nG[n]u8en8n8nT auG-u8nT
.....01100 14002

YeuuVatGIUASHASNLeuSerGlyASNileGluProAlaLeuAlag 84

14351 CATTTAGATGGAGATCCAGGACTGGAGGATACAGACCGGGACTTCGTA 14362

84] 1189TTCGATGGAGGACCTGGCTA 14302

84 IUA^{SP}LYR^{SR}LEU^{SR}SER^{AS}GLY^{LE}THR^{TY}THR^{PH}LEU^{LY}S^{LY} 100

14301 AAAGTTGGGATATTCTTGATGACCGGAGTACTTTATATCCCTTC

101 C G A G A G G A G T T A T A C G T T C A T T A A G A 14252

101 ...SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheI116
...

[illegible]

14201 GAATTAACCTTGGAGTGGATGGAGTTCCATTACTGCCGAAGAATA, AG 14203

116 eGluSerTrpLysGlnValAlaThrGlnGluValSerGlyIleThrValLeu123

[illegible]

14202 AAAATCTTATCTTAGAATTTAGATAAAGAAACCGGCTCATCTTTGTGTA 14153

```
133 healAleu...AsnProfil[el]vsDenVal[Ar]v[er]t[ic]e[al]
```

148

14152 ACATGATTAGTCCGTTATTAAAAATCGAGAAGAGTATTTTGACGGCAA 14103

[illegible]

147 LeuserIleAspHisPheGlyValHisSerProAsnGluSerThrLeuVa 165

14102 GCAAATGAGTCTGAGCCTTGGAAATTAAAGCTCTTCAATCAAAAAAAAAAATCTTTCT 14103

102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

165 lValThrLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeup 182

14052 AATAAGGCTAAATCTCCACCGC
::|::|::|::|::|
::|::|::|::|::|

14003 5'-GCTAAGGCTAAATACTCCAAAGCCATATTTCTTGATATGTTAGTACATC

182 roValPhePheProVal.....HisLysSerGlnArgThrLeuGlnSer 106

1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718

14002 AACATTTATTCTGTACCAATGCACGTTATTGAAAGTATGGCAAGG 13953

197 LysserlePro

127 LysSerLeuPro.....IleAlaSerGlyAlaPheTyrProLysAs 210

13952 TGGACAGACCCCGAGAAATATCGTTGTTAGTGGTCCCTTCATATTAATATTC 12003

CTGAGTGGTTTCATTAATAAATC 13903

210 nIleLysGlnLysGlnTrpIleLysLeuSerLysAsnProHisTyrTyrA 227

13902 TAGAGTATTAAATCAGGACCTTCCTCCCG

18502 1AGAGG111AGAAAGGTTGTTCTTGAAAAGAATAATAAATTATA 13853

```

alignment_scores:
  Quality: 443.00
  Ratio: 1.443
  Percent Similarity: 58.365
  Length: 526
  Gaps: 13
  Percent Identity: 26.046

alignment_block:
  US-09-824-567-2 x AAX13139 ..
  Align seg 1/1 to: AAX13139 from: 1 to: 10996

30 SerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArgSe 46
   :::::::::::::: |||::: ::::::::::::::||
   1 ACTGCCGCTAAACAGAAATTTAAAGTAGTAGTTCACAGAAGATTCCTT 50
   |||

46 rLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuValIysH 63
   |||

51 TGCAGATTATCTTTAGCAACATACATATTAGTTTACGGCACTGAATA 100
   |||

63 isletTyrrGluGlyValGlnGluAsnLeuSerGlyAsnIleGlu 79
   :: |||||::: :: :: :: :: |||

```

```
seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/Na1999.DAT.AAX13139
seq_documentation_block:
ID AAX13139 standard: DNA; 10996 BP.
```

```

101 ATGCGTATGAAGGAATTTATCGTTTAGATGATTAAGCAGCCGCAACCA 150
80 ProAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrThr 96
151 CGAGGGCTAAAGAAAGTGAAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 200
96 rPheLysLeuLys...SerAlaPheTrpSerAsnGlyAspProLeuThr 112
201 AGTGAACACTAAGAGAGCAAAATGGTCCATGGTGATCCTGTCACAG 250
112 laGluAspPheLeuSerTrpLysGlnValAlaThrGlnGluValSer 128
251 CTGCGGATATGTTTATGTTGGCAAGTACAGCAGATCCTCAGACTGGC 300
129 GlyIleTrpAlaPheAlaLeuAsnProIleLysAsnValArgLysIleG 145
301 GCTGAGTATGCTATTTCTTGAATATGTTGAACAGGTCGACAGCATGT 350
145 nGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGlu 162
351 AGCTGGTAAGAAACCAAGTCTCTGAATTAGCATTAAGCAACAGGAGAT 400
162 erThrLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeu 178
401 ACAGCTAGCAATTAACCTGGCAAAACCAACACCATATTTGATTATTTA 450
179 LeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLeuG 195
451 CTGGCTTTCCACTATTTTCCGCGCAACAT.....CAAGCAACCGPAGA 494
195 n.....SerLysSerLeuProIleAlaSerG 204
495 GAAATATGGCAAGATTAACGCTCTCCAGTGAAGAAAGCAGTCATAATG 544
204 lyAlaPheTyrProLysAsnIleLys.....GlnLysGlnTrp 216
545 GTCCGTTTGTCTTACCTAATTTGAAGGGCTGTCAGATACCAATGG 594
217 IleLysLeuSerLysAsnProHisTyrTrpAsnGlnSerGlnValGluTh 233
595 ...ACTTTGGAAAAAATGAACACTATTTGGATTAAGACAAATGTCAAA 641
233 rlysthrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysL 250
642 AGATAAAATCAATTTGATGTTGTCAAGAGGCGCCCAACCGCTTCAACT 591
250 euPheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArg 266
692 TATTCGAAGTGGCCAGGCGAGC.....GATGTTATT 723
267 IleProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSe 283
724 TTAAGCGCGCAATTTGCCAGCAAGATTCGGAAGGATCCAGATTTGTTAT 773
283 rPheAspValAlaGlyThrSerTrpLeuThrPheAsn.....IleAsnL 298
774 CGAAAAAGAACAGCAGCAACAGCTATTTGGAATTTATCAGCGAGATAAAA 823
298 ysPheProLeuAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeu 314
824 ATTCACCGTATAATGATGATTTAGCCAAAGCAATTTTCAGCAGCGATT 873
315 AspLysGluAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAl 331
874 GATCGCAACGCTTAGTAGATAAATTTTAGCGGACGGCTCAGTAGTCGC 923
331 aAspHisLeuLeu.....ProThrAsnIleHis 341
924 TACCGGCTTAATCCAGAGGATGTCGTATTTCGCCGACAGATGACACAG 973
341 erTyrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAla 357
974 ATTTCTGCTGATAAATAAAAAAATTTTGAATATATAGCCGCAAGCA 1023

```

```

358 LysLysLeuPheLysGluAlaLeuGluGlnIleThrAlaLysas 374
1024 AAAGAATATTGGCAAAAGCGAAGAAATTTGGGCATAACAACACTCAA 1073
374 pleuGluHisLeuAsnLeuIlePheProValSerSerSerAlaSerSerL 391
1074 AATGGAT.....ATTGTTGCCGATGATGTCGATTTCTACGAAAA 1111
391 euLeuValGlnLeuIleArgGluGlnTrpLysGluSerLeu...GlyPhe 406
1112 AGTTAGCGGAATATTAAGGTAGCTTTAAGAGTACTTTTAGAGGCATT 1161
407 AlaIleProIleValGlyLysGluPheAlaLeuGlnAlaAspLeuSe 423
1162 GATGTACAGCTGAGCCAGTCGCCGTTTTCAGTTCTCGAATTGATCGAGGTAG 1211
423 rSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaA 440
1212 TCGCGGCGATTTTGAGACAACTTAGGTGGTTGGCTGCGGATTACGAG 1261
440 spPrometalPheLeuThrIlePheAlaTyrProSerGlyValProPro 456
1262 ATCCAGTAGTTTCTTAGAICTTTTGTACTGTATAACTATAATATCGC 1311
457 TyrAlaIleAsnHisLysAspPheLeuGluIleLeuGln.....AsnI 471
1312 GGGCGTTTTCAGTAAAGCTACGATGAGTTGATTGAGGCTTCAGCTAC 1361
471 eGluGlnGluAspHisGlnLysArg...SerGluLeuValSerGlnA 487
1362 AAGAGATGCTTCAGATCCTGAAAAACCTTGGGAAGACATGGTTAAGGCTG 1411
487 laSerLeuTyrLeuGluThrPheHisIleLeuProIleTyrHisasp 503
1412 AGAAGTTATTGTTGGTGAAGAGACCGCTTTAGCACCACTTTATCAAAAG 1461
504 AlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerPr 520
1462 GCGACTGCCCATCTACGCGATTAAGAGTGAAGGCGTGTGCCCCATGG 1511
520 oThrGlyVal...ValAspPheArgTyr 528
1512 TGGTGGCGCACAAATATGATTATAAGTGG 1539

```

seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAS46237

seq_documentation_block:

ID AAS46237 standard; DNA; 12790 BP.

AC AAS46237;

DT 18-DEC-2001 (first entry)

DE DNA encoding novel mar regulated protein (NIMR) #6.

KW mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.

OS Escherichia coli.

PN WO200170776-A2.

PD 27-SEP-2001.

PF 08-MAR-2001; 2001WO-US07478.

PR 10-MAR-2000; 2000US-188362P.

PA (TUFT) TUFTS COLLEGE.

PI Levy SB, Barbosa TV, Alekshun MN;

XX WPI; 2001-602769/68.

DR


```

2693 AGGAAATATTTCACCTTTTAAACATCGGCATATGACAAAGTCTGCGCC 2742
470 snleGluGlnGluGlnAspHisGlnLysArgSerGluLeuValSerGln 486
2743 AGGCATCGACGGAATAACCGTTAAGCGCGTAATGCGGATTACACAGCG 2792
487 AlaSerLeuTyrLeuGluThrPheHisIleIleGluProIleTyr 501
2793 GCAGAAAAATCCCTATGAGCAAGCAGCGATTGCACCAATTTAT 2837

```

seq_name: /sids5/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.ABA03041

seq_documentation_block:

ID ABA03041 standard; DNA; 2944528 BP.

AC ABA03041;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes EGD-e genome sequence.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease; ds.

OS Listeria monocytogenes.

PN W0200177335-A2.

PD 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

PA (INSP) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusnok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;

DR WPI; 2002-010914/01.

PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and
PT related polypeptides

PS Claim 1: SEQ ID No 1; 192pp; French.

CC The present sequence is the genome sequence of *Listeria monocytogenes*
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in *L. monocytogenes* and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Proteins (AB847297-AB850149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of *L.*
CC monocytogenes and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate *L.*
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccine compositions for
CC the treatment or prevention of infections by *L. monocytogenes* and related
CC organisms.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

```

alignment_scores:
  Quality: 470.50      Length: 579
  Ratio: 1.368        Gaps: 20
  Percent similarity: 59.413  Percent identity: 25.043
alignment_block:
US-09-824-567-2 x ABA03041/rev ..
Align seg 1/1 to reverse of: ABA03041 from: 1 to: 2944528
1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
2286215GTGAAATCTAATATTCTTACACTTGGATTACACTATTACTTAAG 2286166
17 rValValLeuGlnGlyCys.....LysG 25
2286165CTTAGCTTGGTAGCGGCGGATCCAAATCCGACAAA 2286116
25 lueSerSerHisSerThrSerArgGlyGluLeuAlaIleAsnIleArg 41
2286115AAGGCTCAGATTAGGAAAGCTTCAGGAGAGCAAGTACTTAACCTTCA 2286066
42 AspGluPro.....ArgSerLeuAspProArgGlnValArgLeuLeuSe 56
2286065GAAAGCGCACTAATCTCTCGACAGACGACAAAGCGGATGCCAAGT 2286016
56 rGluIleSerLeuValHisIleTyrGluGlyLeuValGlnGluAsnA 73
2286015TGGTTTGAACGTTGTAACAAACAAAGAGGCTCTATATGCG..... 2285973
73 snLeuSerGlyAsnIleGluProAlaLeu.....AlaGluAspTyr 86
2285972..CTTGACAAAGATGTTATCTCCGCTGCTGCTGAAGAGCCA 2285925
87 SerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys.....SerAl 102
2285924AAATAGGCGATGACAAACAGTTTACTATCTCAACCTTCGGAAGATGC 2285875
102 aPheTyrSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerT 119
2285874AAATGGTCAACGAGGAGCTCTGTAACATGACATGTTTACTCAT 2285825
119 rLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheAlaLeu 135
2285824GGCGTCTGCAGTTGACCTTAACACTGCTGCAACATATTCTTACTATT 2285775
136 AsnProIleLysAsnValArgLysIleGlnGlyHisLeuSerIleAs 152
2285774GATGCAATCAAAACGCGTGAGATATCTAGTCGCAAGAAAACCTGA 2285725
152 pHisPheGlyValHisSerProAsnGluSerThrLeuValValThrLeuG 169
2285724AGATTAGGAATTAACAGCAGTAGATTATATTAGAGTTACTCTTT 2285675
169 lueSerProThrSerHisPheLeuLysLeuAlaLeuProValPhePhe 185
2285674TAAACCACTGCTTACTTAACCTTTCGCAATTCCTCCCAACTTCTTC 2285625
186 ProValHis.....LysSerGlnArgThrLeuGlnSe 196
2285624CCACTTAACGAAAAATTCGTACGGAAGGCGGAAAATATGCAACAAA 2285575
196 rLysSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIle.... 211
2285574TAGGATAACATCTTCAATGAGACCTTCGAGTTCAAGAGATTGGACTG 2285525
212 ..LysGlnLysGlnTrpIleLysLeuSerLysAsnProHisTyrAsn 227
2285524GACACAAACAAAATGGACTTACGTA...AAAAATGATAAATATTGGAT 2285478
228 GlnSerGlnValGluThrLysThrIleThrIleHisPheIleProAspAl 244

```

2284646CTTTATCAAGTTCCTACTGCATACCTACAAAAGACTACATATAA
516 uGlyValSerProThrGlyValValAspPheArgTyr 528
| : : : : :
2284596GCAGAAAAATCCATTGGT...CCAGANTACACATTAC 2284563

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 04:36:22 ; Search time 87.81 Seconds
(without alignments)
1048.096 Million cell updates/sec

Title: US-09-824-567-2
Perfect score: 2739
Sequence: 1 MRKISVGICIFILLSLVWL.....LSNLGVSPGTGVDFRYAKEN 532

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL19:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_phase:**

10: sp_plant:**

11: sp_rodent:**

12: sp_virus:**

13: sp_vertebrate:**

14: sp_unclassified:**

15: sp_rvirus:**

16: sp_bacteriaph:**

17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2739	100.0	532	16 Q928Y9	Q928Y9 chlamydia p
2	1364	49.8	518	16 Q84201	Q84201 chlamydia t
3	1345	49.1	520	16 Q9PKJ4	Q9PKJ4 chlamydia m
4	1160.5	42.4	527	16 Q928Y8	Q928Y8 chlamydia p
5	755.5	27.6	526	16 Q928Y6	Q928Y6 chlamydia p
6	755.5	27.6	528	16 Q9K244	Q9K244 chlamydia p
7	748	27.3	527	16 Q9PKL9	Q9PKL9 chlamydia m
8	685.5	25.0	529	16 Q84178	Q84178 chlamydia t
9	536.5	19.6	550	16 Q97D46	Q97D46 clostridium
10	509.5	18.6	528	2 Q9K5F7	Q9K5F7 borrelia ga
11	506	18.5	528	2 Q9K5J9	Q9K5J9 borrelia af
12	505.5	18.5	528	2 Q31305	Q31305 borrelia bu
13	505.5	18.5	528	16 Q54584	Q54584 borrelia bu
14	482.5	17.6	435	16 Q928Y7	Q928Y7 chlamydia p
15	479.5	17.5	532	16 Q9AKR0	Q9AKR0 rhizobium m
16	478	17.5	559	16 Q929H6	Q929H6 listeria in

17	476	17.4	550	2 Q93QH8	Q93QH8 lactococcus
18	474.5	17.3	558	2 Q9LAT7	Q9LAT7 listeria mo
19	472	17.2	550	16 Q9CIL3	Q9CIL3 lactococcus
20	471.5	17.2	530	16 Q31315	Q31315 borrelia bu
21	467.5	17.1	514	2 Q53480	Q53480 borrelia co
22	467	17.1	545	16 Q9CIL2	Q9CIL2 lactococcus
23	446.5	16.3	553	16 Q927S4	Q927S4 listeria in
24	446	16.3	551	2 Q9FS06	Q9FS06 bacillus th
25	446	16.3	565	16 Q97D48	Q97D48 clostridium
26	441	16.1	529	2 Q9FJ34	Q9FJ34 enterococcus
27	440.5	16.1	519	2 Q9RMW8	Q9RMW8 bacillus an
28	431	15.7	543	16 Q9KT14	Q9KT14 vibrio chol
29	429.5	15.7	523	2 Q31313	Q31313 borrelia bu
30	429.5	15.7	529	2 Q31303	Q31303 borrelia bu
31	428.5	15.6	529	16 Q50927	Q50927 borrelia bu
32	428.5	15.6	530	16 Q985N6	Q985N6 rhizobium l
33	427.5	15.6	523	2 Q52615	Q52615 borrelia bu
34	425.5	15.5	529	16 Q51307	Q51307 borrelia bu
35	421	15.4	549	16 Q92FA6	Q92FA6 listeria in
36	419	15.3	426	16 Q84141	Q84141 chlamydia t
37	415.5	15.2	541	2 Q52616	Q52616 borrelia bu
38	414.5	15.1	541	16 Q51308	Q51308 borrelia bu
39	412	15.0	428	16 Q9PKP9	Q9PKP9 chlamydia m
40	412	15.0	549	16 Q9C127	Q9C127 lactococcus
41	411.5	15.0	523	2 Q31304	Q31304 borrelia bu
42	409.5	15.0	541	2 Q31306	Q31306 borrelia bu
43	400.5	14.6	547	16 Q83594	Q83594 treponema p
44	390.5	14.3	545	2 Q51643	Q51643 enterococcus
45	379	13.8	522	2 Q06515	Q06515 streptococcus

ALIGNMENTS

RESULT	1
Q928Y9	
ID	Q928Y9
AC	Q928Y9; PRELIMINARY; PRT; 532 AA.
DT	01-MAY-1999 (TREMBL:10, Created)
DT	01-MAY-1999 (TREMBL:10, Last sequence update)
DT	01-DEC-2001 (TREMBL:19, Last annotation update)
DE	OLIGOPEPTIDE BINDING PROTEIN.
GN	OPPA_1 OR CPN0195 OR CP0572.
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).
OC	Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.
OX	NCBI TaxID=83558;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CWL029;
RX	MEDLINE=99206606; PubMed=10192388;
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.,"
RL	Nat. Genet. 21:385-389(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=AR39;
RX	MEDLINE=20150255; PubMed=10684935;
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA	Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA	Eisen J., Fraser C.M.;
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT	pneumoniae AR39.,"
RL	Nucleic Acids Res. 28:1397-1406(2000).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=J138;
RX	MEDLINE=20330349; PubMed=10871362;
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138

RT from Japan and CWL029 from USA.
 RL Nucleic Acids Res. 28:2311-2314 (2000).
 DR EMBL: AE001606; AAD18348.1; -
 DR EMBL: AE002216; AAF38391.1; -
 DR EMBL: AF002545; BAA98405.1; -
 DR TIGR: CF0572; -
 DR InterPro: IPR000914; SBP_bac_5.
 DR Pfam: PF00496; SBP_bac_5; 1.
 KW Complete proteome.
 SQ SEQUENCE 532 AA; 59744 MW; 1CB473D9D46A1579 CRC64;

Query Match 100.0%; Score 2739; DB 16; Length 532;
 Best Local Similarity 100.0%; Pred. No. 1.5e-196;
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISVGICITILLSVLVLOGCKSSHSSTSRGELAINIRDEPRSLDPRQVRLLS 60
 DB 1 MRKISVGICITILLSVLVLOGCKSSHSSTSRGELAINIRDEPRSLDPRQVRLLS 60

QY 61 VKHIYEGVLQVNNLSGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFIESW 120
 DB 61 VKHIYEGVLQVNNLSGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFIESW 120

QY 121 QVATQVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLLA 180
 DB 121 QVATQVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLLA 180

QY 181 LPVFPVHKQSRTLSQSLPIASGAFYPKNIKOKWIKLSKNPHYNSQVETKTIHF 240
 DB 181 LPVFPVHKQSRTLSQSLPIASGAFYPKNIKOKWIKLSKNPHYNSQVETKTIHF 240

QY 241 IPDANTAAKLFNQGKLNWGPWGERIPQETLSNLSQSGHLHSFVAGTSLWTENKFP 300
 DB 241 IPDANTAAKLFNQGKLNWGPWGERIPQETLSNLSQSGHLHSFVAGTSLWTENKFP 300

QY 301 LNNMKREALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERQKQEMAQRAYAKL 360
 DB 301 LNNMKREALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERQKQEMAQRAYAKL 360

QY 361 FKEALELOITAKDLEHNLIFPVSSASSLLVQLIREQWKSLSGFAIPVIGKEFALLQA 420
 DB 361 FKEALELOITAKDLEHNLIFPVSSASSLLVQLIREQWKSLSGFAIPVIGKEFALLQA 420

QY 421 DLSSGNFSLATGWFADPADMAFLTIFAYPSGVPPYAINHKDFLEILQNTQEQDQHR 480
 DB 421 DLSSGNFSLATGWFADPADMAFLTIFAYPSGVPPYAINHKDFLEILQNTQEQDQHR 480

QY 481 SELVQSASLYLETFHIEPIYHDAFOFAMNKKLSNLGVSPGVVDVFRYAKEN 532
 DB 481 SELVQSASLYLETFHIEPIYHDAFOFAMNKKLSNLGVSPGVVDVFRYAKEN 532

RESULT 2
 O84201 PRELIMINARY; PRT; 518 AA.

ID O84201
 AC O84201;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE OLIGOPEPTIDE BINDING PROTEIN.
 GN OPPA_3 OR CT198.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UV-3/CX;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kalnan S., Lammie C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:

RT Chlamydia trachomatis.
 RL Science 282:754-759 (1998).
 DR EMBL: AE001293; AAC67790.1; -
 DR InterPro: IPR000914; SBP_bac_5.
 DR Pfam: PF00496; SBP_bac_5; 1.
 KW Complete proteome.
 SQ SEQUENCE 518 AA; 58858 MW; 8B3AE840831BBEF1 CRC64;

Query Match 49.8%; Score 1364; DB 16; Length 518;
 Best Local Similarity 50.1%; Pred. No. 9.8e-94;
 Matches 269; Conservative 97; Mismatches 147; Indels 24; Gaps 8;

QY 1 MRKISVGICITILLSVLVLOGCKSSHSSTSRGELAINIRDEPRSLDPRQVRLLS 56
 DB 1 MRKISVGICITILLSVLVLOGCKSSHSSTSRGELAINIRDEPRSLDPRQVRLLS 56

QY 57 EISLVKHIEGLVQVNNLSGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAEDF 116
 DB 57 DINLIHLYEGVLQVETP-SGVPFPALESFLESDEKTYTFNLKAFWSNGDPLTAEDFV 115

QY 117 ESKQVATQVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFL 176
 DB 116 RSNVDVLRVASIYSAFUPI-DVAK-----DSGFFAKDDHTLVINLLTPPHFL 165

QY 177 KLALPVFPVH-KSRTLSQSLPIASGAFYPKNIKOKWIKLSKNPHYNSQVETK 235
 DB 166 KLTLVPFVPSHQSRIKKEKSLPISTGAFLEKKDRRLKLEKSPYYNKQVAVQE 225

QY 236 ITHTIPDANTAAKLFNQGKLNWGPWGERIPQETLSNLSQSGHLHSFVAGTSLWTEN 295
 DB 226 ICHIIIPQQAALFNGKLNWGPWGERIPQETLSNLSQSGHLHSFVAGTSLWTEN 285

QY 296 LNNMKREALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERQKQEMAQRAYAK 355
 DB 286 TAKKPFESHKLQALSVLNKEALASLAFV---APAKHLLPAHLHTYPPQSYKQOEAIT 342

QY 356 YAKLFKEALELOITAKDLEHNLIFPVSSASSLLVQLIREQWKSLSGFAIPVIGKEF 415
 DB 343 LAKSLLEALTELMNTLEKYPVLTFTSATSTMSIAQLRQWRSLGTFPICKEY 402

QY 416 ALLQADLSSGNFSLATGWFADPADMAFLTIFAYPSGVPPYAINHKDFLEILQNTQEQ 475
 DB 403 ALLQNDLIGNTFWMSGWGFADPADMAFLTIFAYPSGVPPYAINHKDFLEILQNTQEQ 475

QY 476 DHQKSELVQSASLYLETFHIEPIYHDAFOFAMNKKLSNLGVSPGVVDVFRYAKEN 532
 DB 462 NPKRSALISEASLYTERQNVIEPLYHDFVHYTNNKLSFVRLHPSGLVDMRYAKNS 518

RESULT 3
 Q9PKJ4 PRELIMINARY; PRT; 520 AA.

ID Q9PKJ4
 AC Q9PKJ4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN,
 DE PUTATIVE.
 GN TC0471.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / NIGG;
 RX MEDLINE=20150255; PubMed=10584935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

```

RT Pneumoniae AR39.1;
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002315; AAF39319.1; -.
DR TIGR; TC0471; -.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
KW Complete proteome.
SQ SEQUENCE 520 AA; 59116 MW; 93E9C106CC0C7F58 CRC64;

Query Match 49.1%; Score 1345; DB 16; Length 520;
Best Local Similarity 49.4%; Pred. No. 2.6e-92;
Matches 266; Conservative 103; Mismatches 145; Indels 24; Gaps 9;

QY 1 MKKISVGICITILLSSVVLQCKESS-----HSSTSRGELAINTRDPRSLDPRQVRL 55
DB 1 MKKVSVGIC--LLIALAVATGCKSSKSNKSSNQ--SVSVSKMDPRTFDPREVRL 57
QY 56 SEISLVKHYIEGLVQNNLNIENIPALAEYDSSDGLTYTFKLSAFWSGDLTAEDF 115
DB 58 SDINLHLYEGLVQETP--SGEVPALAESEFTLSEDKTYTFHKLKALNSGLITAHDF 116
QY 116 IESKQVATQEVSGIYAFALNPKNVKKIQEGHLSIDHFGVHSPNESTLVVLTSPSHF 175
DB 117 VRSNDVQLNRITASYSPAFLLPI-----DLSKO--CGFFAKDNHTLVNLTFTPHF 166
QY 176 LKLLALVPFPVH--KSQRTLOSKSLPTASGAFYPNKIKQWIKLSKNPHYNSQVETK 234
DB 167 LKLLTLPVPVPHPHQIRNEAKALPISGACFLCKEKKDKREWKLEKNPYVYKNEQVAT 226
QY 235 TITIHFPDANTAALFNQGLNWQPPGGERIPQETLSNLSQSKGHLHGFVAGTSWLTFF 294
DB 227 ETHIVIPDQATSALEKQGLDWGLPWGHSIQEALATANKRRTPQSGTISGTSWLTFF 286
QY 295 NINKEPLNMKRLALASALKEALVSTIFLGRAKTADHLLPTNIHSEYEHQKQMAORQ 354
DB 287 NTSKVPESHKPLQALSLVNLKEALASPTFV---KPAKHLTPHLLTHTYPEQFTYKQEA 343
QY 355 AVAKLEKEALELQITAKDLEHLNLIFFVSSASSLLVOLIREQWESLGAIPVIGKE 414
DB 344 ILAKTLQEALETDLNMTIKDLKCLPLISATSVNSVNSQAMQMDQWRILGITFPGCKE 403
QY 415 FALQADLSGNSFLATGWFADFADPMAFUTIFAYPSGVPYVAINHKDFLEILQNIQE 474
DB 404 YALLONDLTNTFMGINSWFADFSDPLAFUSVPS--SKGIKPYALQDPLFDQLLSIETE 462
QY 475 ODHOKRSELYSQASLYETHTIEPIYHDAFOFAMNKLNSLGVSPGTGVDFRYAKZN 532
DB 463 KDKRKRISLSEASLYLEKQNIIEPLHYDHFHYAANKLSFVRLPSGLVDMRYAKNS 520

RESULT 4
Q928Y8 PRELIMINARY; PRT; 527 AA.
AC Q928Y8
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE OLIGOPEPTIDE BINDING PROTEIN.
GN OPA.2 OR CPN0196 OR CP0571.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.1";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi P., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AE001606; AAD18349.1; -.
DR EMBL; AE002215; AAF38390.1; -.
DR TIGR; CP0571; -.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
KW Complete proteome.
SQ SEQUENCE 527 AA; 61166 MW; 628F32FB4F13D79E CRC64;

Query Match 42.4%; Score 1160.5; DB 16; Length 527;
Best Local Similarity 43.2%; Pred. No. 1.7e-78;
Matches 225; Conservative 111; Mismatches 176; Indels 9; Gaps

QY 15 SISVLQCKESSHSSTSRGELAINTRDPRSLDPRQVRLSEISLVKHYIEGLVQNNL 74
DB 11 TLWLITSGC---SPSQSKGIFVYVKNKEMPRSLDPGKTRTADQTLMRHLYEGLVEEHSQ 67
QY 75 SGNIEPALAEYDSSDGLTYTFKLSAFWSGDLTAEDFIESKQVATQEVSGIYAF 134
DB 68 NGEIAPALAEYTTISDGTTRYTPKKNILWSGDLTAEDFVSWKEILKEDASSVLYA 127
QY 135 LNPKNVKKIQEGHLSIDHFGVHSPNESTLVVLTSPSHFLKLLALPVPFVHKSORTL 194
DB 128 FLPIKNARAIFDDTESPENLGVRLDKRHLSIQLETPCAHFLHFLPLFPFVHETLRNY 187
QY 195 QS--KSLPIASGAFYPNKIKQWIKLSKNPHYNSQVETKTIHTIFPDANTAALFN 252
DB 188 STSFEEMPITCGAFRPSLEKGLRLHLEKNPMYHNSRVKLHKIIVQFISNANTAALFK 247
QY 253 QGKLNWQGPWGERIPQETLSNLSQSKGHLHGFVAGTSWLTFTNINKFPLNMKRLALAS 312
DB 248 HKLDWQGPWGPGEIPPEISASLHQDDQLFSLPGASTWLLFNLIQKPNNAKRLKALS 307
QY 313 ALDKEALVSTIFLGRAKTADHLLPTNIH--SYEHQKQMAORQAYAKLFKEALELQI 370
DB 308 AIDKMDLTKVYQGLAETFDHLLAPRLYPGTYPYERKRN--ERILEAQOLFEEALDELQ 365
QY 371 TAKDLEHLNLIFFVSSASSLLVOLIREQWESLGAIPVIGKEFALLQADLSGNSFLA 430
DB 366 TREDLEKETLTSTSTSFYGRICQMLREQWKKVLAFTIYVQGEFFTIQNFLEGNYSLT 425
QY 431 TGGWFADFADPMAFUTIFAYPSGVPYVAINHKDFLEILQNIQEQQDKHSELSVQASLY 490
DB 426 VNQWTAFTDPMSYMILFANPGGISTPYHLQDSHFQTLIKTIQEHKKHLNQLITDALY 485
QY 491 LETFTHIEPIYHDAFOFAMNKLNSLGVSPGTGVDFRYAKE 531
DB 486 LEHCHLPELCHPNRLALNKNKFNLFVRRTSDFRIEK 526

RESULT 5
Q928Y6 PRELIMINARY; PRT; 526 AA.
ID Q928Y6

```

QZ8Y6;	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	OLIGOPEPTIDE BINDING PROTEIN.
GN	OPPA_4 OR CPN0198.
OS	Chlamydia pneumoniae (Chlamyidophila pneumoniae).
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamyidophila.
OX	NCBI_TaxID=83558;
ON	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CMLO29;
RC	MEDLINE=99206606; PubMed=10192388;
RX	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RA	*Comparative genomes of Chlamydia pneumoniae and C. trachomatis.;
RL	Nat. Genet. 21:385-399(1999).
RT	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=J138;
RC	MEDLINE=20330349; PubMed=10871362;
RX	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT	*Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT	from Japan and CMLO29 from USA.;
RL	Nucleic Acids Res. 28:2311-2314 (2000).
DR	EMBL; AE001606; AAD18351.1; -
DR	EMBL; AP002545; BAA98408.1; -
DR	InterPro; IPR000914; SBP_bac_5.
DR	Pfam; PF00496; SBP_bac_5; 1.
KW	Complete proteome.
SQ	SEQUENCE 526 AA; 59822 MW; D149730C9E2C18C CRC64;

[illegible]

RESULT	6
Q9K244	
ID	Q9K244
AC	PRELIMINARY; PRT; 528 AA.
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE	PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN.
GN	CF0569.
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX	NCBI_TaxID=83558;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	STRAIN=AR39;
FX	MEDLINE=20150255; PubMed=10684935;
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA	Linhier K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA	Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.
RA	Eisen J., Fraser C.M.;
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT	pneumoniae AR39.";
RL	Nucleic Acids Res. 28:1397-1406(2000).
DR	EMBL; AE002215; AAF38388.1; "
DR	TIGR; CP0569; "
DR	InterPro; IPR000914; Snp_bac_5.
DR	Pfam; PF00496; Snp_bac_5; l.
SQ	SEQUENCE 528 AA; 60082 MW; F2722C0CC9E2DB11 CRG64;

Query Match	27.68:	Score	755.5:	DB	16:	Length	528:
Best Local Similarity	33.98:	Pred.	No. 3.2e-48:				
Matches	177:	Conservative	106:	Mismatches	216:	Indels	23:
Gaps							
QY	13	LLLSVVVLOGCKESHSHSTGRGLAINIRDEPSRLSDQPROVRLSELISLVKHYEGLVOEN	72		:	:	:
Db	17	LLFLLLTSSCKQKQEPGLK-HLVITAMSHDIADUDPRNAYLSRDSALAKALYELGRLT	75		:	:	:
QY	73	NLSNTEPIALAEYSLSDGLTYTFKLKSAFWSNGDPLTAEDFIESKOVATQVSGIYA	132		:	:	:
Db	76	D-QGIALALAEVSTLSDKHVYTFKLSPSWSDGTLTADEFKSIKQLYFEFSPSH	133		:	:	:
QY	133	FALNPILKNVYKIQEHLISIDHFGVHSNPESTLWLTBSPSTSHFLKLLALPVPFFVHKSR	192		:	:	:
Db	134	TLGVINKSSAIHNAQKSLETGLQAKDOLLTVLTBQPPYFLTIARPVFSVPHUTR	193		:	:	:
QY	193	TLQSKSLP----IAGAFYPKNIKQWIKLSKNPNHYNOSQVETKTIITHFIPDANTAA	248		:	:	:
Db	194	ESYKTKTPPSYISNGPFVKKHEHNTYLLEKNPHYDHESVKLDRTLKIIPDSTAT	253		:	:	:
QY	249	KLFNOGKLNWQGPWGERIPQETLSNLOSKGHLHSDVAGTSLWTFNFKNPFLNNMKLE	308		:	:	:
Db	254	KLFKSKSIDWIGSPWAPSINED-QKVLQSKELITVSSTTLIYNLOKPLIQNKALRK	312		:	:	:
QY	309	ALASALDKREALVSTIFLGRAKTADHLPTNTHSYPEHOKEMAOQAQYAKKLKEAL	368		:	:	:
Db	313	ATAHAIDRKSLRLVPSG--QEAVLTPPNLSQINLOKEISTEERTKARAYFEAKETL	370		:	:	:
QY	369	QITAKDLEHNLPIPVSSASSLLVQLIREQWKSGLGPAITVGKEFALLQADLSGGNFS	428		:	:	:
Db	371	--SEKELAEALSILYIDSSNSSIIAQEIQLAKDVLGLKIKIQMEYHCFLKKRQGGDF	428		:	:	:
QY	429	LATGQWADADPMALRTIAYPGVPPIYAINHKDFELQNI---EQEDHOKRSSELVS	485		:	:	:
Db	429	IATGQWIAEYVSPVAFSLISILGNPRDLTQW--RNSDYEXTLKLPHAYKENLKAEMI-	485		:	:	:
QY	486	QASLYLTFTHIEPIYHDAQAFQANKKLSNLGVSPTGVVDPR	527		:	:	:
Db	486	-----IEETPPIPLVHGKYIYATHPKIQMTGSLSLIGHTDUL	522		:	:	:

RESULT 7

Q9PKL9 PRELIMINARY; PRT; 527 AA.
 AC Q9PKL9
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN.
 GN TC0446.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / NIGG;
 RX MEDLINE=20150255; PubMed=10684935;
 RA White D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae Ak39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL; AE002313; AAF39300.1; -;
 DR TIGR; TC0446; -;
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 2.
 KW Complete proteome.
 SQ SEQUENCE 527 AA; 60452 MW; 510E8C6B49BF437 CRC64;

Query Match 27.3%; Score 748; DB 16; Length 527;
 Best Local Similarity 35.3%; Pred. No. 1.2e-47;
 Matches 185; Conservative 96; Mismatches 217; Indels 26; Gaps 12;

QY 14 LSLSVVLOGKESHSSTSRGEALINIRDEPRSLDPRVRLSEISLVKHIVYEGVLQENN 73
 DB 14 LSLTGLTSCYHKEE--PKDVLRTAICHDPMSLDPQVFLSKOVSVIKALYEGVLREND 71
 QY 74 LSGNTEPALAEADYSLSDGLTYTFKLSAFWSNGDPLTAEDFTESMKQVATQVSGIYAF 133
 DB 72 --GSHLALAEYSLSDRCVYTFKLTFWHNGDLVTAEDFEESIKQYLFHEVDNVALR 129
 QY 134 ALNPKKNVRKIQEGHLSIDHFGVHSPNESTLVVLTESPTSHFLKLLALPVFPVHKSSRT 190
 DB 130 LLALIKNSHAVLAGDIPVENLGVRLDHTLEITLHPSSHFTLTHPVFPVHSLRE 189
 QY 191 -QRTLOSLSPIAS--GAFYPPKNIKOKWIKLSKNPHYNQSOVETKTITIHFIPTANTAA 248
 DB 190 YRNRSKRSPLTISNGPFIIRCYEPQNFLLDKNPFYHQKNVSLDAYRLQIVPDIHTAV 249
 QY 249 KLFNOGKLNWQGPWGERIPQETLSNLSQSGHLSHSDVAGTSMWLTNFKPLNNMKLR 308
 DB 250 QLFCKKYVDLVGLPSSSFPLEQKNL--SQDFLYDYPVNLCTVLCFNVNHPKLPDPSLR 308
 QY 309 ALASALDKALYSTIFLGRKTAADHLLPTNHSYPEHQEQMAQQAQAKLFEALDEL 368
 DB 309 ALSLAIDRETLK--LACKGSIATSFVHPSLSKNPLDVLQ--KERISLANNALAEAL-- 363
 QY 369 QITAKDLEHLNIFPVSSASSLLVLIREQWKESLGFAIPVIGKERALLQADLSGNSF 428
 DB 364 TVPQELKXITLIVPIESIVLRVAVQVIRQLFDVLGFKISTLGLVHSPDKRSRGEFS 423
 QY 429 LATGGWFADFPMAFLTIFAYSGVPPYPAI---NHKDFLEILQN--IEGEQDHQKRSL 483
 DB 424 LSTGNWVADYQAKAFSLGNGTKYKDYQVIDWQNEQYTDIVSRLLVEDSTDLQILAE- 482
 QY 484 VSOASLYLETHFIETPIYHDAFQFAMNKKLSNLGVSPTGVVDPR 527
 DB 483 ----QLLKESPLI-PLYHLDYAYAKHPKVSNLQTSLSGEIDLK 521

RESULT 8

O84178 PRELIMINARY; PRT; 529 AA.
 AC O84178;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE OLIGOPEPTIDE BINDING PROTEIN PERMEASE.
 GN OPPA_2 OR Cti175.
 OS Chlamydia trachomatis
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kallam S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 DR EMBL; AE001291; AAC67766.1; -;
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 2.
 KW Complete proteome.
 SQ SEQUENCE 529 AA; 60331 MW; 2D546BD1DFC3F786 CRC64;

Query Match 25.0%; Score 685.5; DB 16; Length 529;
 Best Local Similarity 34.4%; Pred. No. 5.5e-43;
 Matches 181; Conservative 89; Mismatches 227; Indels 29; Gaps 10;

QY 14 LSLSVVLOGKESHSSTSRGEALINIRDEPRSLDPRVRLSEISLVKHIVYEGVLQENN 73
 DB 14 VSLAFLTSCYHKEE--PKDVLRTAICHDPMSLDPQVFLSKOVSVIKALYEGVLREND 71
 QY 74 LSGNTEPALAEADYSLSDGLTYTFKLSAFWSNGDPLTAEDFTESMKQVATQVSGIYAF 133
 DB 72 AA--FQLALAEYHQSDGCVYTFKLTFWHNGDLVTAEDFEESIKQYLFHEVDNVALR 129
 QY 134 ALNPKKNVRKIQEGHLSIDHFGVHSPNESTLVVLTESPTSHFLKLLALPVFPVHKSSRT 193
 DB 130 SLALIKNSHAVITGALPVEDIGVRLNAKTLEIVLENPPYFLEILAHVPVFPVHTSLRE 189
 QY 194 L-----QSKSLPIAS--GAFYPPKNIKOKWIKLSKNPHYNQSOVETKTITIHFIPTANTAA 247
 DB 190 YRKDRNRKRVPIISNGPFAIQCEYEPQVRYLLINKNPLYHAKHVDLLNSVCLQIVPDIHTA 249
 QY 248 AKLFNOGKLNWQGPWGERIPQETLSNLSQSGHLSHSDVAGTSMWLTNFKPLNNMKLR 307
 DB 250 MGLFQKNHIDLVGLPSSSFPLEQKNL--PREKLFDPVLSVLCFNIHQTPANNPSLR 308
 QY 308 EALASALDKALYSTIFLGRKTAADHLLPTNHSYPEHQEQMAQQAQAKLFEALDEL 367
 DB 309 TALSIAINRETLK--LACKGSIATSFVHPSLSKNPLDVLQ--KERISLANNALAEAL-- 365
 QY 368 LQITAKDLEHLNIFPVSSASSLLVLIREQWKESLGFAIPVIGKERALLQADLSGNSF 427
 DB 366 L--SQEDLEKITLIVPIESIVLRVAVQVIRQLFDVLGFKISTLGLVHSPDKRSRGEF 423
 QY 428 SLATGGWFADFPMAFLTIFAYSGVPPYPAIHNKDFLET-----LQNIQEQDHQKRSL 481
 DB 424 SLATGNWVADYQAKAFSLGNGTKYKDYQVIDWQNEQYTDIVSRLLVEDSTDLQILAE- 476
 QY 482 ELVSOASLYLETHFIETPIYHDAFQFAMNKKLSNLGVSPTGVVDPR 527
 DB 477 DLQLMAEQLLLKESPLIPLYHLDYAYAKHPKVSNLQTSLSGEIDLK 522

RESULT 9
 Q97D46

ID Q97D46 PRELIMINARY; PRT; 550 AA.
 AC Q97D46;
 DT 01-OCT-2001 (TremBLrel. 18, Created)
 DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE OLIGOPETIDE ABC TRANSPORTER, PERIPLASMIC SUBSTRATE-BINDING COMPONENT.
 GN CAC3634.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Ometchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Ghu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL: AE007859; AK81557.1;
 DR InterPro: IPR000914; SBP_bac.5.
 DR Pfam: PF00496; SBP_bac.5;
 DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
 DR Complete proteome.
 KW
 SQ SEQUENCE 550 AA; 61190 MW; 3A7A48AD99C6C051 CRC64;

Query Match 19.6%; Score 536.5; DB 16; Length 550;
 Best Local Similarity 26.8%; Pred. No. 8.1e-32;
 Matches 153; Conservative 121; Mismatches 232; Indels 65; Gaps 15;
 QY 2 RKISVGICITLLSL-SVVLGGCKESHSTGRGE--LAINIRDEPSLDPQPROVRLLSRI 58
 Db 4 RKTKLSAVLSAALLAGCGSSSTSTGTEQKYVNLGADPTIDPLGNNSEVG 63
 QY 59 SLVKHYELGLQENNLSTEPALAEYSLSDGLTYTFKL-KSAFWSNGDPLTARDTE 117
 Db 64 TWIENAFGLV-DINKKEKVPVGVASSWDISADNLTYTFHLRKNKAKWSGDKPVKARDEF 122
 QY 118 SNKQVATQEVSGITAFALNPKIKWRKIQEHLSDHFGVHSPNESTLVLTLESPTSHFLK 177
 Db 123 AKWRALAPETASDAYOLLYLKNGEAYNNGKASGDVGVKAIDDTYTKVNLAPTPYFLS 182
 QY 178 LLALPVPFPPVHKSORTLOSKSLP-----IASGAPYPNKIKQKWLKSNPHYVNSQOV 231
 Db 183 LTAPTYMPLREDIVSKDNKNAKONYVNSGPFYMTDMKATMTFSKNPNYNNKNTI 242
 QY 232 ETKTITHFTPDANTAALFNQGLNQGPPWGERIPQETLSLQSKGHLHSFDVAGTWS 291
 Db 243 KINSITYMLAQESSATAFTSQVD-----INDLIPAVOKSLTQKGAKAYPYGTGF 297
 QY 292 LTFNIN-----KFLNNKILREALASALDKALYSTIFLGRKATDHLPTNI-- 339
 Db 298 FDINGDKDSANGAEITTKLNPKYREALNLAVDRESIVKNYTKGEPATSFVPSIKL 357
 QY 340 -----HSYPEHOKEMAQROAYAKLFKEA-----LEELQITAKDLEHLNIFP 383
 Db 358 PNGTKFNKNDYP--AKGVKK-----AKQLLAAGYDYGKGPSSQIMYNE----- 402
 QY 384 VSSASSLLVQILREQWKESLGFPAIPVIGKEFALLQADLSSGNFSLATGWFADFADPMA 443
 Db 403 --GSNNQVQALQDMYKKNLINFILQSVKRVQDLNLTQOQYQICRASHIADYNDPMT 460
 QY 444 FLTIFAYPSGVPPIYAINHKDFLETLQNIQEOHQRSELVSOA--SLYLETFFHIEPIYH 502
 Db 461 FMDLFTVDSGNNNGSYSNPEYDALIKDAKTNDADKRIDDMKEADAAMDRLPI--PIYE 519
 QY 503 DAFQFAMNKKLSNLGVSPTGVDFR--YAKE 531

Db 520 YTNVVEVKPYVDLHKLSPGLFVFNNTYIKK 550
 RESULT 10
 Q9K5F7
 ID Q9K5F7 PRELIMINARY; PRT; 528 AA.
 AC Q9K5F7;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE SURFACE ANTIGEN.
 GN P58.
 OS Borrelia garinii.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=29519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PBI;
 RA Wilske B., Habermann C., Fingerle V., Hillenbrand B.,
 RA Jauris-Helpe S., Lehnert G., Pradel I., Roessler D.,
 RA Schulte-Spechtel U.;
 RT "An improved recombinant IgG immunoblot for serodiagnosis of Lyme
 borreliosis";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ250013; CAB96045.1;
 DR HSSP: P06202; IB52.
 DR InterPro: IPR000914; SBP_bac.5.
 DR Pfam: PF00496; SBP_bac.5;
 SQ SEQUENCE 528 AA; 60562 MW; AA6F2B802CBC46FB CRC64;

Query Match 18.6%; Score 509.5; DB 2; Length 528;
 Best Local Similarity 28.6%; Pred. No. 7.9e-30;
 Matches 138; Conservative 104; Mismatches 20; Indels 39; Gaps 14;
 QY 38 INIRDEPSLDPQPROVRLLSLSESLVKHYELGLQENNLSTEPALAEYSLSDGLTYTF 97
 Db 32 ISLSEPSLSDQLADNDVNGSKMIDTMRGLITGDPNTGNKPLGAKSWDISPDGTWTF 91
 QY 98 KLK-SAFWSNGDPLTAEDEFESKQVATQEVSGIYA-FALNPKIKWRKIQEHLSDHFG 155
 Db 92 TLREKIWSGDVAITAEIRKSYRLINKETGNSYAEAMVKSTIKNGQYFDQGVSDSELG 151
 QY 156 VHPNESTLVLTLESPTSHFLKLLALPVF--FPVHKSORTLOSKSLP---IASGAPYPN 210
 Db 152 IRAIDEKTELTLESPKPYFDMLVHQSFIPPIHIAEKYQSWISPENIVTSGPKLKE 211
 QY 211 IKQKWLKSNPHYVNSQOVETITTHIFPDANTAALFNQGLNQGPPWGERIPOE 270
 Db 212 RIPNEKYVVEKNDKYNSNOVEQETFTYTTNDSSTAYKMYENKELD---AIFGS-IPD 267
 QY 271 TSLNLSQKGLHLSFDVAGTWSLTFTNFKPLNNKILREALASALDKALYSTIFLGRAT 330
 Db 268 LKLDKLSDDYSSAVNAIYFAFNTPYIKPLDNKVRKALTALDRETLTYKV-LDNGTT 326
 QY 331 ADHLAPTNIHSPHOKEMAQROAYAK--KLFEALFEELQITAKDLEHLNIFPV---- 384
 Db 327 PTRRIAPNFSSY-----SYAKNLEFNPEITAKTLIAEAGYPNGN-GFPILK 373
 QY 385 --SSASSLLVQILREQWKESLGFPAIPVIGKEFALLQADLSSGNFSLATGWFADFADPM 442
 Db 374 YNTSEAHKKICEFIQWKKILINDVELENEETWYTLNTRNSNGVEIARAGWIGDYADPL 433
 QY 443 AFLTIFAYPSGVPPIYAINH---KDFLETLQNIQEOHQRSELVSOA--SLYLETFFHIE 498
 Db 434 TFLSIFT--QGYTFQSSHNSSPYENELIKKLSDELDPKRODILRKAETIIEKDPPIA 491
 QY 499 PIY 501
 Db 492 PIY 494
 RESULT 11


```

Q9K5J9
ID Q9K5J9 PRELIMINARY; PRT; 528 AA.
AC Q9K5J9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SURFACE ANTIGEN.
GN P58.
OS Borrelia afzelii.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
ON NCBI_TaxID=29518;
RX MEDLINE=98240240; PubMed=9579077;
RC STRAIN=PKO.
RP SEQUENCE FROM N.A.
RA Wilske B., Habermann C., Fingerle V., Hillenbrand B.,
RA Jauris-Helpe S., Lehnert G., Pradel I., Roessler D.,
RA Schulte-Spechtel U.,
RT "An improved recombinant IgG immunoblot for serodiagnosis of Lyme
RT borreliosis."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ250012; CAB6044.1; -.
DR HSP; P06202; IBS2.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PIR; P06202; IBS2.
SQ SEQUENCE 528 AA; 60509 MW; E2D048F7290A7A16 CRC64;

Query Match 18.5%; Score 506; DB 2; Length 528;
Best Local Similarity 28.0%; Pred. No. 1.4e-29;
Matches 139; Conservative 106; Mismatches 210; Indels 42; Gaps 14;

QY 24 KESSHSSTSRGELAINRDRSPQVRLSEISLVKHYEGLVGNLSNIEPALA 83
DB 21 KEKKEKGV---FKISGAPSSLDQLADNDVNGSKMIDTFKGLITGDPNTGNKPGLA 77

QY 84 EYLSLSDGLTYFLK-SAFWSNGDPLTAEDFIESWKQVATQVSGIYA-FALNPIKNV 141
DB 78 KSWIDSPDQVYVTLREKIISDGVAITAEGIRKSYRLINKETGSGNYSVAKSTIKNG 137

QY 142 RKIQSGHLSIDHFGVHSPNESTLVVLESPTSHFLKLALVFPV-PPVHKSQRTLSQSKL 199
DB 138 QKVFQGVSDSELGIRAIIDEKTEILTESPKPYFDMLVHQSFIPV-PPVHKSQRTLSQSKL 199

QY 200 P---IASGAFYKNIKOWIKLSKNPHYNSQVETKTITHTPDANTAAKLFNGKLNMQGPPWGERIP 268
DB 198 PENIVTSQVSDSELGIRAIIDEKTEILTESPKPYFDMLVHQSFIPV-PPVHKSQRTLSQSKL 199

QY 257 NQQPPWGERIPQETLSNLSQKHLHSFVAGTSMVTNFKPLNNMKLREALASALDK 316
DB 258 D----AIPSAIPDLKRLSDYSSAVNAIFYAFNTHIKPLDNVAVKRAITLAIDR 313

QY 317 EALVSTIFLGRKTAADHLLPTNIHSGYPEHOKQMAQRAYAK--KLFEALBELQITAKD 374
DB 314 EFLTYKV-LDGTTPRTTPNPFSSY-----SYAKNLELFPETAKTLAEAG 360

QY 375 LEHNLIFPV-----SSASLLVLQIREQWKSGLGFAIPVKGFEALLQADLSGNGFS 428
DB 361 YPNGN-GFPILKLYNTSEAHKICEFQNGKKNLIDVLENEEWITVNTKANGYEIARAGWIGDYAD 431

QY 429 LATGWAFADPMAFLTFAYPSGVPPYAINH---KDFLEILQIEQODHOKRSELYSQA-SLYLETHI 496
DB 420 IARAGWIGDYADPLTFLSILTH--GYTFSSSHNSYSPYNEILKKSDELDPKIRQDILR 477

QY 486 QA-SLYLETHIETPIY 501
DB 478 KAEIIEKDFPIAPIY 494

RESULT 12
ID O31305 PRELIMINARY; PRT; 528 AA.
AC O31305;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)

```


Db 216 AIIHLNRGKVDWQFQWGHGIPWEL--HKOSQYHYTYTVEGAFWLCINTKSPHLNDLQN 273
 QY 307 REALASALKEALVSTIFLGRKADHLLPTNIHSYPERHOKQMAQOAYAKLKFKEALE 366
 Db 274 RRLATCIDKRSIIEEALQGGTQQAETL-----SRGAPQPNQYKK----- 313
 QY 367 ELQITAKDLEHLNLIFFPVSSASSLLVQLIREQWKESLGFAPIVGKEFALLQADLSSGN 426
 Db 314 --QKPLTPQEKLVLYPSDILRCORIAEILKEQWK-AAGIDILLEGLEYHLFVNRKRVQD 370
 QY 427 FSLATGGWADPADPMALFTITFAPVSGVPPYAINHKDFLEILQNTQEQDHOKRSELVSQ 486
 Db 371 YAIAT-----QTGVATYPG-----ANLISE 390
 QY 487 ASLYLETFHIIPIYHDAFOFAMNKLNLGNSPTGVDFRY 528
 Db 391 EDKLLQNFI--PIYLSYDYLTDQFIEGVYINASGAVDLKY 431

RESULT 15
 Q9AKR0
 ID Q9AKR0 PRELIMINARY; PRT; 532 AA.
 AC Q9AKR0;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE OLIGOPEPTIDE ABC TRANSPORTER (PUTATIVE OLIGOPEPTIDE UPTAKE ABC
 DE TRANSPORTER PERIPLASMIC SOLUTE-BINDING PROTEIN).
 GN OPPA.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GR4;
 RA Nogales J.; Olivares J.; Sanjuan J.;
 RT "Oligopeptide transport systems in Sinorhizobium meliloti.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021; PLASMID=PSYMB (MEGAPLASMID 2);
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorhoeiter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL; AJ296269; CAC33589.1;
 DR EMBL; AL603645; CAC49261.1;
 DR HSSP; P06202; 1B52.
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 1.
 KW Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 532 AA; 59004 MW; FA0A382E4027D691 CRC64;

Query Match 17.5%; Score 479.5; DB 16; Length 532;
 Best Local Similarity 27.1%; Pred. No. 1.4e-27;
 Matches 133; Conservative 103; Mismatches 221; Indels 33; Gaps 11;

QY 43 EPRSLDPRQVRLISLKVHIEGLVQENNLGNIPEALAEYSLSDGLTYFFKLK- 101
 Db 37 EPQTLDQAHTSINIEFFILKDLYGLT- IYDAAGKIVGAAETNELSDGTVYTFKLAD 95
 QY 102 AFWSNGDPTAEDFTESWKQVATQEVSGIYAFALNP IKNRKIOBGHLSDHFGVHSPNE 161
 Db 96 AKWSDGSPVTAEDFAFSRRVEDKTAAREYANILPIKNAEKVNKGEVPDQLGVKAVDE 155
 QY 162 STLWVLTSPSTSHLKLALPVPFPHKQSQ-----RTLQSKSLPIASGAFYPNIKQKW 216
 Db 156 KTLVLTERTPTFFLELLAHQTALPVSKASVEKNAGDFVKPCVMVSNAGFKLTAHYPNDS 215

QY 217 IKLSKNPHYVNSOVERTKITTHFTPDANTAAKLFNKGKLNWQPPWGERIPQETLSNLQ 276
 Db 216 LTVKNTNINWDAANVKLDKVIYFIDQQAASVRRFEAKEMD-----LAYNFSADQIERLR 270
 QY 277 SK--GHLHSFDVAGTSMWTFNINKFPLNNMKLREALASALDKALYSTIFLGRKADHLL 334
 Db 271 TSYGBOVHVSPTLATYYAFDTROEPYNDVRRALSMVDRDFLAKEIYSGQLPSYSM 330
 QY 335 LPTNIHSYDPHQOEMA-----OROQYAKKLFEALEELQITAKDLEHLNLIFFPVSSAS 389
 Db 331 VPPGIESYGDPAKADPADMSQLDREDKAIELMKEA-----GYGEGKPLNIEIRYNINPN 385
 QY 390 -SLLVQLIREQWKESLGFAPV-----VCKEFAFALLQADLSSGNFSLATGGWADPADPM 444
 Db 386 HERVATAVADMWKNITGAKVSLVNLVDVSSHVAYLQ-----EGGKFNVARAGWADYDAENF 442
 QY 445 LTIFAYPSGVPPYA-INHKDFLEILQNTQEQDHOKRSELVSQASLYLETHIIEPIYHD 503
 Db 443 LALSLSNTKTFNYGHFENAEFDALMKKSYBEQDPAARSKIMHEAETLLMKSQPIAPFLTQ 502
 QY 504 AFQFAMNKKL 513
 Db 503 ADUMLVSEV 512

Search completed: July 26, 2002, 04:41:41
 Job time: 319 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 03:36:37 ; Search time 50.88 seconds

(without alignments)
1004.708 Million cell updates/sec

Title: US-09-824-567-2

Sequence: 1 MKRISVGCITITLLSYVL.....LSNLGVSPTGVDERAKEN 532

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 28338 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length DB	ID	Description
1	2739	100.0	532	A72107	peptide ABC transp
2	2739	100.0	532	C86515	oligopeptide bindi
3	1364	49.8	518	B71545	probable oligopept
4	1364	49.1	520	B81699	peptide ABC transp
5	1160.5	42.4	527	B72107	peptide ABC transp
6	1160.5	42.4	527	D86515	oligopeptide bindi
7	755.5	27.6	526	D72107	oligopeptide bindi
8	755.5	27.6	526	F86515	oligopeptide bindi
9	755.5	27.6	526	A81564	peptide ABC transp
10	748	27.3	527	H81701	peptide ABC transp
11	685.5	25.0	529	F71547	probable oligopept
12	617.5	22.5	545	A38447	oligopeptide ABC t
13	536.5	19.6	550	B97346	oligopeptide ABC t
14	505.5	18.5	528	H70140	oligopeptide ABC t
15	489.5	17.9	549	G69618	di-peptide ABC tran
16	482.5	17.6	435	E86515	oligopeptide ABC t
17	482.5	17.6	435	C72107	oligopeptide ABC t
18	479.5	17.5	532	E95949	peptide ABC transp
19	478	17.5	559	AH1719	probable oligopept
20	474.5	17.3	553	B98207	phenomone ABC tran
21	474.5	17.3	553	AG3079	hypothetical prote
22	474.5	17.3	558	AD1349	phenomone ABC tran
23	472	17.2	550	G86667	hypothetical prote
24	471.5	17.2	530	D70218	oligopeptide ABC t
25	467	17.1	545	H86667	hypothetical prote
26	446.5	16.3	553	AD1771	di-peptide ABC tran
27	446	16.3	565	H97345	oligopeptide ABC t
28	438.5	16.0	544	D64882	periplasmic oligop
29	438.5	16.0	553	A11395	di-peptide ABC tran

30	437.5	16.0	544	2	G90867	probable transport
31	437.5	16.0	544	2	B85751	probable transport
32	431	15.7	543	2	B82242	oligopeptide ABC t
33	428.5	15.6	529	2	B70211	oligopeptide ABC t
34	425.5	15.5	529	2	G70140	oligopeptide ABC t
35	421	15.4	549	2	A11457	ABC transporter ol
36	419	15.3	426	2	D71552	probable periplasm
37	418	15.2	538	2	AD0285	probable periplasm
38	417	15.2	545	2	A10285	periplasmic oligop
39	415	15.2	543	1	F64871	oligopeptide-bindt
40	414.5	15.1	541	2	A70141	oligopeptide ABC t
41	412	15.0	428	2	A81705	peptide ABC transp
42	412	15.0	543	2	G90846	hypothetical prote
43	412	15.0	543	2	F85704	hypothetical prote
44	412	15.0	549	2	C86592	peptide-binding pr
45	411.5	15.0	537	2	AB0660	periplasmic murein

ALIGNMENTS

RESULT 1
A72107
peptide ABC transporter, periplasmic peptide-binding protein, probable CP0572 [import
C:/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:/Date: 23-Apr-1999 #sequence-revision 23-Apr-1999 #text-change 11-May-2000
C:/Accession: A72107, AB1562
R:/Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:/Reference number: A72000; MUID:99206606
A:/Accession: A72107
A:/Molecule type: DNA
A:/Residues: 1-532 <RNA>
A:/Cross-references: GB:AE001606; GB:AE001363; NID:94376464; PIDN:AA018348.1; PID:9437
A:/Experimental source: strain CW029
R:/Head, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C./Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salde
Nucleic Acids Res. 28, 1397-1406, 2000
A:/Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR35
A:/Reference number: AB1500; MUID:20150255
A:/Accession: AB1562
A:/Molecule type: DNA
A:/Residues: 1-532 <REA>
A:/Cross-references: GB:AE002216; GB:AE002161; NID:97189484; PIDN:AA038391.1; PID:9716
A:/Experimental source: strain AR39, HL cells
C:/Genetics:
A:/Gene: oppA1; CP0572
C:/Superfamily: dipeptide transport protein

Query Match 100.0%; Score 2739; DB 2; Length 532;
Best local Similarity 100.0%; Pred. No. 1.4e-187;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRISVGCITITLLSYVLCKESSSSSTSGEALNIRDEPSLPDQVRLSEISL 60
DB 1 MKRISVGCITITLLSYVLCKESSSSSTSGEALNIRDEPSLPDQVRLSEISL 60
QY 61 VKHIEGLVQENNLNIGNEPALADYSLSSDGLTYFKLSAFMSNGDPLTAEDFIESNK 120
DB 61 VKHIEGLVQENNLNIGNEPALADYSLSSDGLTYFKLSAFMSNGDPLTAEDFIESNK 120
QY 121 QVATQVSGITVAFALNPVKVNRKIQEGHSLIDHFGVSPNESTLVLTLESPSHFLKLA 180
DB 121 QVATQVSGITVAFALNPVKVNRKIQEGHSLIDHFGVSPNESTLVLTLESPSHFLKLA 180
QY 181 LPVFFVNRQRTQSKSLPLASGATPKNIKOKWILSNPHYVQSOVETKTTTHF 240
DB 181 LPVFFVNRQRTQSKSLPLASGATPKNIKOKWILSNPHYVQSOVETKTTTHF 240
QY 241 IPDANTAAKLINQCKLMWQGPWGERIPQETLSNQSGHLSHDVAGTSMLTNNINRFP 300

Db 241 IPDANTAAALFNQGLNMQGPPWGERIPQETLSNLSQKGLHSPDVAGTSMWLTAFNINPKFP 300
QY 301 LNNMKLRALASALDKREALVSTIFLGRKATADHLPTNINHSYPEHOKQMAQAQAAKLL 360
Db 301 LNNMKLRALASALDKREALVSTIFLGRKATADHLPTNINHSYPEHOKQMAQAQAAKLL 360
QY 361 FKAALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420
Db 361 FKAALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420
QY 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480
Db 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480
QY 481 SELVSOASLYETETHTIEPIYHDAFQFAMNKKLSNLGVSPTGVDFRYAKEN 532
Db 481 SELVSOASLYETETHTIEPIYHDAFQFAMNKKLSNLGVSPTGVDFRYAKEN 532

RESULT 2

C:oligopeptide binding protein [imported] - Chlamydomonas reinhardtii (strain J138)
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: C86515
R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: C86515
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-532 <STO>
A:Cross-references: GB:BA000008; NID:98978568; PID:BA098405.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: oppa_1

Query Match 100.0%; Score 2739; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.4e-187;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRISVGICITLLISLVLOGCKESSHSTSRGELAINIRBEPRLDPROYRLSISL 60
Db 1 MKRISVGICITLLISLVLOGCKESSHSTSRGELAINIRBEPRLDPROYRLSISL 60
QY 61 VKHLYBGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFTESMK 120
Db 61 VKHLYBGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFTESMK 120
QY 121 QVATQVSGITAFALNPKNRKIOEGHLSIDHFGVHSFNESITLVLTLESTSHFLKLA 180
Db 121 QVATQVSGITAFALNPKNRKIOEGHLSIDHFGVHSFNESITLVLTLESTSHFLKLA 180
QY 181 LPVFFVHKSORTLOSLSPLASGAFYPPKNIKOKOMIKLSKNPHYVNOSEVETKTITIF 240
Db 181 LPVFFVHKSORTLOSLSPLASGAFYPPKNIKOKOMIKLSKNPHYVNOSEVETKTITIF 240
QY 241 IPDANTAAALFNQGLNMQGPPWGERIPQETLSNLSQKGLHSPDVAGTSMWLTAFNINPKFP 300
Db 241 IPDANTAAALFNQGLNMQGPPWGERIPQETLSNLSQKGLHSPDVAGTSMWLTAFNINPKFP 300
QY 301 LNNMKLRALASALDKREALVSTIFLGRKATADHLPTNINHSYPEHOKQMAQAQAAKLL 360
Db 301 LNNMKLRALASALDKREALVSTIFLGRKATADHLPTNINHSYPEHOKQMAQAQAAKLL 360
QY 361 FKAALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420
Db 361 FKAALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420
QY 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480
Db 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480

Db 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480
QY 481 SELVSOASLYETETHTIEPIYHDAFQFAMNKKLSNLGVSPTGVDFRYAKEN 532
Db 481 SELVSOASLYETETHTIEPIYHDAFQFAMNKKLSNLGVSPTGVDFRYAKEN 532

RESULT 3

B71545

probable oligopeptide binding protein - Chlamydia trachomatis (serotype D, strain W
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Jan-2000
C:Accession: B71545
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Arevild, L.; Miel
Science 282, 754-759, 1998
A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia
A:Reference number: A71570; MUID:99000809
A:Accession: B71545
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-518 <ARN>
A:Cross-references: GB:AE001293; GB:AE001273; NID:93328597; PID:AA067790.1; PID:93;
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: oppa_3
C:Superfamily: dipeptide transport protein

Query Match 49.8%; Score 1364; DB 2; Length 518;
Best Local Similarity 50.1%; Pred. No. 1.8e-89;
Matches 269; Conservative 97; Mismatches 147; Indels 24; Gaps 8;

QY 1 MKRISVGICITLLISLVLOGCKESSHSTSRGELAINIRBEPRLDPROYRLSISL 56
Db 1 MKRISVGICITLLISLVLOGCKESSHSTSRGELAINIRBEPRLDPROYRLSISL 56
QY 57 EISLVKHYEGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFT 116
Db 57 EISLVKHYEGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFT 116
QY 116 RSMNDVLONRVASTISYFAFLPT-DVAK-----DSGFPAKDHTLVINILTPPHFL 165
Db 116 RSMNDVLONRVASTISYFAFLPT-DVAK-----DSGFPAKDHTLVINILTPPHFL 165
QY 166 KLTLLPYFVYHSHQHOIRKEKSLPSTGAFPLKEKKDRMFKLESPYYNKKQYAVOE 225
Db 166 KLTLLPYFVYHSHQHOIRKEKSLPSTGAFPLKEKKDRMFKLESPYYNKKQYAVOE 225
QY 226 ITHIFDANTAAALFNQGLNMQGPPWGERIPQETLSNLSQKGLHSPDVAGTSMWLTAFNINPKFP 285
Db 226 ITHIFDANTAAALFNQGLNMQGPPWGERIPQETLSNLSQKGLHSPDVAGTSMWLTAFNINPKFP 285
QY 286 TAKRPFHSHKRLQALSLVANKALASLAFV---KPAKHLLPAHLRTPEQPSYKQQAAT 342
Db 286 TAKRPFHSHKRLQALSLVANKALASLAFV---KPAKHLLPAHLRTPEQPSYKQQAAT 342
QY 342 YAKLFEALAEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 402
Db 342 YAKLFEALAEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 402
QY 402 ALLQADLSSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 461
Db 402 ALLQADLSSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 461
QY 461 FKAALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420
Db 461 FKAALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420
QY 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480
Db 421 DLSGNSFLATGGMFADFPADPMALFTIFAYPSGVPYAINKDKLELTONIEQEDHQKR 480

Query Match 49.1%; Score 1345; DB 2; Length 520;
Best Local Similarity 49.4%; Pred. No. 4e-88;
Matches 266; Conservative 103; Mismatches 145; Indels 24; Gaps 9

```

RESULT      5
B72107
peptide ABC transporter, periplasmic peptide-binding protein, probable CP0571 [imported]
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 11-May-2000
C:Accession: B72107; C81564
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; M0ID:99206606
A:Accession: B72107
A:Status: preliminary
A:molecule type: DNA
A:Residues: 1-527 <ARN>

```

Query Match	42.48	Score 1160.5	DB 2	Length 527
Best Local Similarity	43.28	Pred. No. 6e-75		
Matches 225	Conservative 111	Mismatches 176	Indels 9	Gaps 4

```

Query Match 42.48; Score 160.5; DB 2; Length 527;
Best Local Similarity 43.28; Pred. No. 6e-75;
Matches 225; Conservative 111; Mismatches 176; Indels 9; Gaps 4;

QY 15 SLVVLQGCSSSSSTSRGELAINIRDRPRLDRROYRLISELSIVKHIEGVLQENLL 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 TMLTITSC---SPSQSSKGIFFVNNKEMPRSRDLGKRLIADDTLKRHLYESGLVEIHSQ 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 75 SCNIEPALAEDYSLSSDGLTYTFPKLSAFNSGDELTAEDFIESKQVATOEVSQIYAPA 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 NGEIKPALAESYITISEDGIRYTRFKIKNILMSGDPLTAQDFVSSWKMLKEDASSYLYA 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 135 LNPRTNKKLOEGLSLDHGCVHSPRESLTVYLTLESPISEKLKLLPVFFVHKSQRTL 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 FLPIPNARALPDTESEPNLGVALLDKRHLLETOLETQCAHRLHFLPLIFFVPEHTLRNY 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 195 QS--KSLPIASGAFVPYNIKOKOMIKLSKNPHYNOSOVETKITTHIFPDANTAKLFN 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 STSEFMRPTICGAFRVSLSLEKGRILHLKNNMYHNKSRVLLHKIIYQFISNMTALILFK 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 253 QGKLWMOGPMWGERIQOETLSLNOSKGLHSPFVAGVSWLTFENKFPINNNKREALAS 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 HKKLIDWOGPMWGPRIPEPTEISASLHODDQLFSLPGASPTWTLFNIOKKPMNNKKLRKALSTL 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 313 ALDREALVSTFLPGRAKTAADHLPTNHH--SYPRHOKQENAOQAVALKEALELELOT 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 AILDMDLTKVYVYGLAEPTDHLPRLYPGYIPKRRQN--ERKLEAQOLFEEALDELQW 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 371 TAKDLEHLNIFPVSSASSLWQIREOMKESIGFPIYVGEKAPALQADISSGNSLA 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 TREDLEKETLTFSTFSYGRICQMLRQMKRYVALKTIPIVGOEFYTLQKNLEGNSTLT 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 431 TGGFAPAFAPMAFLTIPIYBSGVPPVIAINHKOFLLEINIOEDODHOKRSELYSOASLY 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 VNMQTAFLDPMYSYMLITAFANPGISPYHLADSHFOCLLIKIQEHKKHRLNQDILTEALDY 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 491 LETFHIEPIYHDAFOFANKKLSNLGVSPTGVADVRYAKE 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 LEHCHILEPLCHNRIALNNKINIKNFULPYRRKSRDRFTLEK 526
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6
D86515
oligopeptide binding protein [imported] - Chlamydomophila pneumoniae (strain J138)
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: D86515
R;Shital, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T., et al.
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A;Reference numbers: A86491, M01D:20350349
A;Accession: D86515
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-527 <STO>

[illegible]

1. 0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.11 0.12 0.13 0.14 0.15 0.16 0.17 0.18 0.19 0.20 0.21 0.22 0.23 0.24 0.25 0.26 0.27 0.28 0.29 0.30 0.31 0.32 0.33 0.34 0.35 0.36 0.37 0.38 0.39 0.40 0.41 0.42 0.43 0.44 0.45 0.46 0.47 0.48 0.49 0.50 0.51 0.52 0.53 0.54 0.55 0.56 0.57 0.58 0.59 0.60 0.61 0.62 0.63 0.64 0.65 0.66 0.67 0.68 0.69 0.70 0.71 0.72 0.73 0.74 0.75 0.76 0.77 0.78 0.79 0.80 0.81 0.82 0.83 0.84 0.85 0.86 0.87 0.88 0.89 0.90 0.91 0.92 0.93 0.94 0.95 0.96 0.97 0.98 0.99 1.00


```

Db 192 ESKKGTIPSTYISNGPFVLLKKHEHONYLLEKNPHYHDESVKLDRTVLTKIIPDSTAT 251
QY 249 KLFNOGLKMOGPPGGERIPQETLSNLOSGHLSHDVAGSWLTENKPEPLNNKLE 308
Db 252 KLFKSKSDIWGSPWSPAPISNED-QKVLSEKILTYVSSTLLIYMLQKPLIONKALK 310
QY 309 ALASALDKKALVSTIFLGAKTADHLPTNHSYPEHOKQEMARQAVAKKLEAEEL 368
Db 311 ALHAIDRKSLILVPSG--QEAVTLPNLSQMLQKEISTEERQTKANAFQEKETL 368
QY 369 QITAKDLEHLNLIFFVSSASSLLVOLIREQKRESLGAIPYKGERALLQADLSSGNFS 428
Db 369 --SEKLAELSLVPIIDSSNSIIAQEIQRQLDKTLKIKIQMEYHCFIKRRQDDF 426
QY 429 LATGWFADPADMAFLITFAVPSGVPYAINHKKDELELIONT---EEDODHOKSELS 485
Db 427 IATGGMIAEVSYPVAFSLITGNPRDITQW--RNSDEKTELEKYLPHAYKENLKRKREMI- 483
QY 486 QASLYLEFPHILEPIYHDAFOFAMNKKLSNLGVSPGVVDFR 527
Db 484 -----IEEPTPIPLVHGKYIYAIHPKIONTFGLGHTDLK 520

RESULT 9
A:Species: Chlamydia pneumoniae
C:Accession: A81564
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81564
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <REA>
A:Cross-references: GB:AE002215; GB:AE002161; NID:g7189472; PIDN:AAF38388.1; PID:g718948
A:Experimental source: strain AR39, HL cells
C:Gene(s):
A:Gene: CP0566
C:Superfamily: dipeptide transport protein

```

```

Query Match 27.6%; Score 755.5; DB 2; Length 528;
Best Local Similarity 33.9%; Pred. No. 4,8e-46;
Matches 177; Conservative 106; Mismatches 216; Indels 23; Gaps 9;

```

```

QY 13 LLSLVYLOGCKESSHSTSGELAINIRDEPSLDPROVRLSLSLVKHIEGLVOEN 72
Db 17 LFLPLTLSSCKQOEPLGK-HLYVIMSHDLADLDPNNAYLSRSLAKALVEGLREI 75
QY 73 NLSGTEPALADYSLSDGLTYTFKLKSAFMSNGDPLTAEDFIESKQAVATQVSGIYA 132
Db 76 D--QGIALALASLYLSDKHVYTFKLKPSVMSDGPPLTADFEKSIKOLFEESPSIH 133
QY 133 PALNPKNKRIQEGHLSIDHFGVHSPNSSTLVLTLESPTSHFKLALVPEPVKSR 192
Db 134 TLLGVINKSSAIIHNAOKSLETLGIAKDDTLVITLLEPPYFTLLTARVESPSVHHLR 193
QY 193 TLQSKSLP-----IASGAFYRKNIKOKWKIKLSKNPHYVNSOVSFTKTIHFPDANTAA 248
Db 194 ESKKGTIPSTYISNGPFVLLKKHEHONYLLEKNPHYHDESVKLDRTVLTKIIPDSTAT 253
QY 249 KLFNOGLKMOGPPGGERIPQETLSNLOSGHLSHDVAGSWLTENKPEPLNNKLE 308
Db 254 KLFKSKSDIWGSPWSPAPISNED-QKVLSEKILTYVSSTLLIYMLQKPLIONKALK 312
QY 309 ALASALDKKALVSTIFLGAKTADHLPTNHSYPEHOKQEMARQAVAKKLEAEEL 368
Db 311 ALHAIDRKSLILVPSG--QEAVTLPNLSQMLQKEISTEERQTKANAFQEKETL 370

```

```

QY 369 QITAKDLEHLNLIFFVSSASSLLVOLIREQKRESLGAIPYKGERALLQADLSSGNFS 428
Db 371 --SEKLAELSLVPIIDSSNSIIAQEIQRQLDKTLKIKIQMEYHCFIKRRQDDF 428
QY 429 LATGWFADPADMAFLITFAVPSGVPYAINHKKDELELIONT---EEDODHOKSELS 485
Db 429 IATGGMIAEVSYPVAFSLITGNPRDITQW--RNSDEKTELEKYLPHAYKENLKRKREMI- 485
QY 486 QASLYLEFPHILEPIYHDAFOFAMNKKLSNLGVSPGVVDFR 527
Db 486 -----IEEPTPIPLVHGKYIYAIHPKIONTFGLGHTDLK 522

RESULT 10
A:Species: Chlamydia muridarum, Chlamydia trachomatis MOPn
C:Accession: H81701
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: H81701
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <TEU>
A:Cross-references: GB:AE002313; GB:AE002160; NID:g7190484; PIDN:AAF39300.1; PID:g719
A:Experimental source: strain N199 (MOPn)
C:Gene(s):
A:Gene: TC0446
C:Superfamily: dipeptide transport protein

```

```

Query Match 27.3%; Score 748; DB 2; Length 527;
Best Local Similarity 35.3%; Pred. No. 1.6e-45;
Matches 185; Conservative 96; Mismatches 217; Indels 26; Gaps 12;

```

```

QY 14 LLSLVYLOGCKESSHSTSGELAINIRDEPSLDPROVRLSLSLVKHIEGLVOEN 73
Db 14 LSLFTGLSLCHQKEEN--LNRIIRVAICHDPWSDLPDROVRLSDVCAKALVEGLREND 71
QY 74 LSGNTEPALADYSLSDGLTYTFKLKSAFMSNGDPLTAEDFIESKQAVATQVSGIYA 133
Db 72 --GSLHLALAEVSLSDQRCVYTFFLKTFPHNNDVLTALDEESIKQTYLHEVDNALR 129
QY 134 ALNPKNKRIQEGHLSIDHFGVHSPNSSTLVLTLESPTSHFKLALVPEPVKSR 190
Db 130 LLAALKSHAVLKGDIPEVNLGVRAIDHETLEITLHSPSHFETLTHVFPVNASLRE 189
QY 191 QRTLOSKSLPIAS-GAFYRKNIKOKWKIKLSKNPHYVNSOVSFTKTIHFPDANTAA 248
Db 190 YRRNRSKSLPIISNGPFIIRCYEPQNLILLDKNPFYHDKNVSIDAVLQIVPDIHNAV 249
QY 249 KLFNOGLKMOGPPGGERIPQETLSNLOSGHLSHDVAGSWLTENKPEPLNNKLE 308
Db 250 QLFQKRYVDLVAGPWSSPFLBEQKNL--SQDFLYLDVYLVNCTYLCVNNHNPFLNPSIRA 308
QY 309 ALASALDKKALVSTIFLGAKTADHLPTNHSYPEHOKQEMARQAVAKKLEAEEL 368
Db 309 ALSLAIDRETLK--LAGGSLATSFVAPSLSKKPLDVLSQ--KERISLAKNYLAELK-- 363
QY 369 QITAKDLEHLNLIFFVSSASSLLVOLIREQKRESLGAIPYKGERALLQADLSSGNFS 428
Db 364 TVROBELKATITLITLPIESTLVRAVVOELRQOLDVYGFKISTGLFYSFLDKRSKGEFS 423
QY 429 LATGWFADPADMAFLITFAVPSGVPYAI--NKKDLELIONT---EEDODHOKSELS 483
Db 424 LSTGNVADYQOAKAFSLILNGTKYKDYIVDMQKQETDVLVSLVLESDTDLIQILAE-- 482
QY 484 VSOASLYLEFPHILEPIYHDAFOFAMNKKLSNLGVSPGVVDFR 527

```


Db 483 ----QLLKESPLI-PLVHLDYAYAKHPKVSNLQISSIGEDLK 521

RESULT 11

F71547

Probable oligopeptide binding protein - Chlamydia trachomatis (serotype D, strain UW3/CX)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence-revision 13-Sep-1998 #text-change 21-Jan-2000
C:Accession: F71547

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac

A:Reference number: A71570; MUID:99000809

A:Accession: F71547

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-529 <ARN>

A:Cross-references: GB:AE001291; GB:AE001273; NID:93328573; PIDN:AA67766.1; PID:9332857

A:Experimental source: serotype D, strain UW-3/CX

C:Genetics:

A:Gene: oppA_2

C:Superfamily: dipeptide transport protein

Query Match 25.0%; Score 685.5; DB 2; Length 529;

Best Local Similarity 34.4%; Pred. No. 4,8e-41;

Matches 181; Conservative 89; Mismatches 227; Indels 29; Gaps 10;

Db 14 LSLAVYLQCKESSHSTRGELAINIDERSLDPPQVLLSEILVHYIEGIGENN 73

14 VSLAFLGLSCYHKKEE--PKDVLRIALICHDMPSLDPROVFLSDSVYKALIEGLVKE 71

74 LSGNIEPALAEDYSSDGLTYTFKLSAFNSNDPLTADPTESMKOVAYQESGIYAF 133

72 AA--FQALAEFRHQSDCCVTFPLKTFMSNGDVYTADEFESIKOYFREIDNPSLR 129

134 ALNPKVKKIOEGHLSIDHEGVSNSSTVLTLESPTSHFLKALAPVFPYKSCQT 193

130 SLAIKNSHAVLIGALPVEDIGVRALNAKLEIVLNPPYVLETLAHVFPVHTSLRE 189

194 L-----QSKSLPIAS-GAFYKRNKIKOKMILKSNHYVNOGVEYTKTITHTFPDANPA 247

190 YKDKRKRRKVPFLISNGFPALQCEPQRYVLLINKNPLYAKHADVLLNSVCCQIYPLDHTA 249

248 AKLFNGCKLNMOGPPWGERIPQETLSLQSKGHLSEFDAGISWLTENKFPILNNKKR 307

250 MQLFORNNHIDVLGIPWSSSFLERQRL--PREKLPDYPLVSCSVLFCNIDQTLNPSLR 308

308 EALASALDEALVSTIFLGRAKTADHLPTNHSYPERHOKOMARQAVAKKLFKEALRE 367

309 TALSLAINETELK--LAGKCSATSFVHPQLSQIPATTLISQ--DERIALAKGILTALKT 365

368 LQTTADLEHLILFPVSSASSLLVOLLREQMKESLQFAPIYKGFALLQADSSGNF 427

366 L--SQEDLEKILITIPYESVCLAVVQLRQGLFVGLFISTLGLIEYHGFIDKRSRGEF 423

428 STATGWRADPADMAFLTFPAVSGVPYATNKHDELEI-----LQNEDEQDHQKRS 481

424 STATGWTADYHQAFLSVLG-----NGTRYKDQLINMOKYNTINAVQLLIQSSS 476

482 ELVQSASILETFPIHPIPIYHAFQFAMKKLSNIGSPGVVDFR 527

477 DLQMAEQDLLKESPLIPLVHLDYAYAKQPRVSDLTQSSRGEIDLK 522

RESULT 12

A38447

oligopeptide ABC transport system substrate-binding protein oppA precursor - Bacillus su

N:Alternate names: oppA homolog; sporulation initiation protein spo0A

C:Species: Bacillus subtilis

C:Date: 12-Jul-1991 #sequence-revision 12-Jul-1991 #text-change 20-Jun-2000

C:Accession: A38447; S15230; G69668

R:Rudner, D.Z.; LeDeaux, J.R.; Iretton, K.; Grossman, A.D.

J. Bacteriol. 173, 1388-1398, 1991

A:Title: The spo0K locus of Bacillus subtilis is homologous to the oligopeptide perm

A:Reference number: A38447; MUID:91139580

A:Accession: A38447

A:Molecule type: DNA

A:Residues: 1-545 <RND>

A:Cross-references: GB:M57689; NID:q143602; PIDN:AAA62687.1; PID:q143603

R:Perigo, M.; Higgins, C.E.; Pearce, S.R.; Gallagher, M.P.; Hoch, J.A.

Mol. Microbiol. 5, 173-185, 1991

A:Title: The oligopeptide transport system of Bacillus subtilis plays a role in the i

A:Reference number: S15230; MUID:91194547

A:Accession: S15230

A:Molecule type: DNA

A:Residues: 1-25, 'T', 27-194, 'E', 196-339, 'M', 341-545 <PER>

A:Cross-references: EMBL:X56347; NID:940004; PIDN:CA93787.1; PID:940005

R:Kunstl, F.; Ogasawara, N.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Bouilliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehlich, S.D.; Emmer, P.T.; Ertian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landino

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parto, V.; Pohl, T.M.; Portele

A:Authors: Schleich, S.; Schreier, R.; Scoffone, F.; Seliguchi, J.; Sekowska, A.; Se

t, K.; Tanaka, M.; Tamakoshi, A.; Tanaka, T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchaya

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: 669668

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-545 <RND>

A:Cross-references: GB:299110; GB:AL009126; NID:92633472; PIDN:CA813000.1; PID:926334

A:Experimental source: strain 168

A:Note: (Initiation of sporulation, competence development)

C:Comment: This operon of five genes is homologous to the oligopeptide permease opero

C:Genetics:

A:Gene: oppA; spo0A

C:Superfamily: dipeptide transport protein

C:Keywords: blocked amino end; lipoprotein; membrane protein; oligopeptide transport;

F.1-20/domain; signal sequence #status predicted <SIC>

F.21/Binding site: sn-2,3-diacylglycerol (Cys) (covalent) #status predicted

F.21/Modified site: fatty acylated amino end (Cys) (in mature form) #status predicted

Query Match 22.5%; Score 617.5; DB 2; Length 545;

Best Local Similarity 29.4%; Pred. No. 3,6e-36;

Matches 159; Conservative 107; Mismatches 249; Indels 25; Gaps 13;

Db 10 ITLLSLVVLQCKESSHSTSRGE-----LAINRDEPRSLDPROVRLSEISLV 61

8 VTLMFLFTLVLSACFGGSGSGNGKDKSGKTTLLINIKTEPSSLHGLANDSVSGYI 67

62 KHIYELVQENNLNINIEPALAEDYSSDGLTYTFKLSAF--MSNGDPLTADPTESMK 120

68 KQFFBELTR--INADDEPEGMAKIEFSKDKGTYFTTRIDGKMSNDPVTADDEFYAK 126

121 QVAIVQVSGIYAFALNPPIKRYKIOEGHLSIDHFGVSPNSSTLVLTLESPTSHFLKLA 180

127 WALDPNNSQYAYQLYITKGAEPANATGKSLDVAVKVNDKLTAKLNLNPPYFTEJLA 186

181 LPVFPVHKSGRTLSQSLPIASGAFYKRN--IKOKW-----IKLSKNHYVNOGVEY 233

187 FTYVPIPK--KIAEKKNKWTNAGDDYVNGFPKATAMKHSSTLEKNDQWDDKDKYL 245

234 KTTTHFLPDANTAKLFGNGCKLNMOGPPWGERIPQETLSLQSKGHLSEFDAGISWLT 293

246 KTIQVMTNNNNTELKRFQAGELDVAAGMPLGQ--LPTESLPLTKKDSLVPEYIAVYTK 304

QY 294 FNINKEPLNNMKLRALASALDEKALVSTIFLGRAKTAADHLPTNHSYPEHOKQEMAOR 353
 Db 305 FNTFAKFLDWNVRKRLALYSLDRQSYVANYQGEQIPAAVAAVPTPKGFEENKGYEFDKN 364
 QY 354 QA-YAKLFKEALEELQIT-AKDEHLNLTFFVSSASSSLVQLIREOMKSLGFALPIY 411
 Db 365 DVKTAFLYELKGLKEMGLSKASDLPKRLSYN-TDAHKAKINQVEMKKMKLVDELVD 423
 QY 412 GKFEALLQDLSSGNLSLATGNGFADPADPMALITFAVPS-GVPPYAINHKEDELELQN 470
 Db 424 NSENNYIDKLHSDQDIGRMGWLGDENDPINLELFRKNGNNDTGMENPERKLLNG 483
 QY 471 IEEDODHOKSELSVQA-SLYLEFHEITPEYHDAQVQAMNKKLSNGVSPGVDPFRA 529
 Db 484 SOTETOKTRAKELKAGIFIDEMP-VAPIYTTDTWQDENLKGVIIMPQTEGYFRNA 542

RESULT 13

897346
 oligopeptide ABC transporter, periplasmic substrate-binding component CAC3634 [imported]
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: B97346
 R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: B97346
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-550 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AKR81557.1; PID:g15026735; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3634
 C:Superfamily: dipeptide transport protein

Query Match 19.6%; Score 536.5; DB 2; Length 550;
 Best Local Similarity 26.8%; Pred. No. 2.2e-30;
 Matches 153; Conservative 121; Mismatches 232; Indels 65; Gaps 15;

QY 2 KRISVIGITITLST-SVYLQCKKESHSTSRGCE-LAINIDEPERSIDPROVALLEI 58
 Db 4 RLITLAVLSAALASILLACGSSSSSTGTEQVSYNLGADQITDGLNNVSGG 63
 QY 59 SLVHIYEGVQENNLSGNIEPALAEDYSLSDDLITTFKL-KSAFWSNGDLTAEDPIE 117
 Db 64 TVIEMAEGLV-DINKNEKYVPGVASSWDISADNLTYTEFLRNKAKMSDGRVAKADPEF 122
 QY 118 SMKQVATQEVSGIYAFALNPINKVAKIOEGHLSIDHGVHSPNESTLVVLTLESPTSHFLK 177
 Db 123 AMKRALPETAISYAYQLIYLNGBANNGKASDKDGVAKAIDDTYLVKLEAPTFELS 182
 QY 178 LLAIPVEFVHKSORLTQSKSLP-----IASGAFYKNIKOKMIXISKPHYVNOY 231
 Db 183 LTAPEYMPREDIYSKDKMAKKNYNSNPFMTMTMKKAKMTSKNPNWNNKNTI 242
 QY 232 ETKTITIHFIIPANTAAKLNQGLNMGPPMGERIPORTLSNLSQSKGHLSEFVACTSW 291
 Db 243 KLSNITTYMLAQESSATIAFTSGVD-----INDLIPAKQSLIOKDKAKIYYIGTFY 297
 QY 292 LITENIN-----KFLNNMKLRALASALDKALVSTIFLGRAKTAADHLPTNI-- 339
 Db 298 FDLNVDGDSANGAITTTKLNPKVRLALAVDRSIVKNTVGTGGEKPAISFVPSIKL 357
 QY 340 -----HSPFHQKQEMAQORAVAKLKEA-----LEELQITADLEHLNLTIP 383
 Db 358 PNRKTFKNKDYP-AKGDVAK-----AKQLAEAGYPPGKGPFSMQLMYNE----- 402
 QY 384 VSSASSLLVQIREQKESIGFAIPIVGKEFALLQDLSSGNLSLATGNGFADPADPMA 443

Db 403 --GSNNQVQALQDMYKKNLINFILQSVKRYQVLDNLTKOQYQICRASWADNDPWT 460
 QY 444 FLTIFAYSSGVPPYAINHKKDEILQNLQEOBDHOKSELSVQA-SLYLEFHEITPEYH 502
 Db 461 FMDLFTVDSGNNNSGVSNPEYDALIKDAKTTMDAKRIDMKHRAEDAAARDLPVLPYE 519
 QY 503 DAFQFAMKKNLSNGVSPGVDPFR--YAKE 531
 Db 520 YTNVEYKPYKDLHNSPLGFEYFNNTYIKR 550

RESULT 14

870140
 oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (oppa-2) homo1
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 20-Aug-1999
 C:Accession: H70140
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: H70140
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-528 <KLE>
 A:Cross-references: GB:AE001139; GB:AE000783; NID:g2688225; PIDN:ACC6707.1; PID:g268
 A:Experimental source: strain B31
 C:Superfamily: dipeptide transport protein

Query Match 18.5%; Score 505.5; DB 2; Length 528;
 Best Local Similarity 28.0%; Pred. No. 3.3e-28;
 Matches 136; Conservative 109; Mismatches 197; Indels 43; Gaps 15;

QY 38 INTRDEPSLDPROVRLSEISLVKHIYEGVQENNLSGNIEPALAEDYSLSDDLITTF 97
 Db 32 ISIGABPSSLDQLAEDNVASKMIDTFRGIVGDPNKGKPGLAGWISSDGIVTFY 91
 QY 98 KKSASF-WSNGDPLTAEDFIESMKQVATQEVSGIYA-FALNPINKVAKIOEGHLSIDHFG 155
 Db 92 NLREKTIWSDGALITABGIRKSTILRLIMKETGSKYVEMVKVINGKQYDQVDTSEIG 151
 QY 156 VHSPESTLVLTSSPTSHFLKILA-----LPVFFVHKSORLTQSKSLP---IASGAFY 208
 Db 152 IRAIDEKLTLETLESPPKPYFLDMLVHQSPFIV-PVHYTEKYQGNWTSPENNVTSQPFKL 209
 QY 209 KNIKQKQIKIKSKPHYVYNOYETITIHFIIPDANTAAKLFNQGKLMWGPMPGERIP 268
 Db 210 KERITENEYVEKKNKKYDSENEVELEIFETTYTNSYAAKWEENELD--AIRGS-IP 265
 QY 269 QETLSNLSQSKGHLSEFVACTSWLTFNINKEPLNNMKLRALASALDKALVSTIFLGRA 328
 Db 266 PRLINLKLKRSYDSSAVNATFYFAFNTHIKPFLDNVAKIRALTLADRETLTYKY-LVNG 324
 QY 329 KTADELFTNHSYPEHOKQEMAQORAYAK--KLKALBELQITAKDEHLNLTIPV-- 384
 Db 325 TTPFRATPNPSSY-----SYAKSLLEFNFETIKTLAAGYIPNGN-GFPLIK 371
 QY 385 -----SSGASSLLVQIREQKESLGFALPIVGKEFALLQDLSSGNLSLATGNGFADPAD 440
 Db 372 LKYVNTNANKKICEFIONQMKNNLIDVLENEBMTYTLNTRKNGEYELARAGWIGYAD 431
 QY 441 PMALITFAFVPSGVPPYA---INHKDEFLQNLQEOBDHOKSELSVQA-SLYLEFHEI 496
 Db 432 PLTFLTSIFT--QGYTFQSSHNATSNPEYELIKKSDLELDTIRQDILQAEIIIEKDFP 489
 QY 497 IEPTY 501
 Db 490 IAPY 494

RESULT 15

G69618

dipeptide ABC transporter (dipeptide-binding protein) - *Bacillus subtilis*C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: G69618; S16651

R:Runt, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azeredo, V.; Berton, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chd Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Geller, lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koster, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauder, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.K.; Portetelle Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schaefer, R.; Scorfione, F.; Sekiguchi, J.; Sekowska, A.; Serot T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: G69618

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1549 <run>

A:Cross-references: GB:299110; GB:AL009126; NID:92633472; PIDN:CA813153.1; PID:92633650

A:Experimental source: strain 168

K.Machopoulos, C.; Mueller, J.P.; Slack, F.J.; Murphy, C.G.; Patankar, S.; Bukusoglu, G Mol. Microbiol. 5, 1903-1913, 1991

A:Title: A *Bacillus subtilis* dipeptide transport system expressed early during sporulation

A:Reference number: S16646; MUID:92114768

A:Accession: S16651

A:Molecule type: DNA

A:Residues: 7-271, 'K', 273-344, 'P', 346-549 <MAT>

A:Cross-references: EMBL:X56678; NID:948802

A:Note: the authors translated the codon GTG for residue 10 as B, AAA for residue 37 as Leu, GTT for residue 182 as Glu, and ACA for residue 183 as Val

C:Genetics:

A:Gene: dppE, dclA

C:Superfamily: dipeptide transport protein

Query Match 17.98; Score 489.5; DB 2; Length 549;

Best local similarity 26.08; Pred. No. 4.9e-27;

Matches 143; Conservative 116; Mismatches 244; Indels 47; Gaps 13;

QY 12 ILLSLVVLQCC-----KSSH-SSTSGE--LAIRNDEPRSDPPROVALSEISL- 60

DB 18 LALGLSFLMGTCTANEGAGKESHDKAKTSGEKVLYNNENEPSPDPTG--FNNVSMQ 75

QY 61 -VKHIYEGVQENNLGNINIPALADYSLSSDGLGYTFKLR-SAFWSNGDPLTADFTES 118

DB 76 PLNNIMBGLTRFLCK-DHEEPAPAEKWSYKDNKTYFTIRNAKWTNGDPTACDFEYA 134

QY 119 WKQVATQVSGIYAFALNPIKNVRIQEGHLSIDHGVHSPNESTLVLTLESPTSHFLK 178

DB 135 WKRMIDPKKAGSAFLGTFLEGGAYNSGKGRKDDVATKDRITLEVTLTAPQKFLSY 194

QY 179 LALVFPFVRK-----SQTLOSLSPLASGAFYPKNIKOKWIKSKNPHYNSOYET 233

DB 195 VSNPAPYFVNEKVDKNDPKWFAESDTPYGNPFKLTBMKHDSITMKSSTYWDKDFYKL 254

QY 234 KTIITIRIPANTAAKLFNQGKLNMGQPPGGERIPQETLSNLSQSKGLHSFVAGTSMLT 293

DB 255 DKVWAWVSDRNDYQMFQSGELD-----TAYVPAELSDQLDQDNVNIYDQAGLYFYR 308

QY 294 FNIKPEPLNMLKREALASALDKREALYSTIFIGRAKTADHLLPTNHSYPEHOKQENA-- 351

DB 309 FNVNMEPFQENIRKAPMAAYDQSEIVKYVTKNNKTA-HAFVSPGTQPDGKDFRAGG 367

QY 352 -----QROAVAKLFEKALEELQITANDLEHLNIPVSSAS-----LLVQLREQWKE 402

DB 368 DLIKPNSKAKQLEKQWKE-----ENVYKPLAITLSTPEPKKIAELIQKLN 419

QY 403 SLGFALPIVKGEPALLOADSSGNFSLATGWFAPADPMAFLTFAYPSGVPAINIK 462

DB 420 SLGVYKLANKEWVLEDDKALKFQFSOSFLPDYADPISFLFAFOTGSMNRGMANK 479

QY 463 FLELLQNLIEDODHOKRSELSQASLYETFHIEPIYHDAPQFANKKLSNLGVSPPG 522

DB 480 EVDQILQKANKADEKTRFSLMQALELLINAPLIPYFYNOVHLQNDQVGIVRHPVG 539

QY 523 VYDERAKEN 532

DB 540 YIDLKWDKN 549

Search completed: July 26, 2002, 04:39:59

Job time: 3802 sec

Enterococcus faecalis
Enterococcus faecalis
B. burgdorferi anti
B. burgdorferi anti
Enterococcus faecalis
B. burgdorferi anti
Enterococcus faecalis
Enterococcus faecalis
Enterococcus faecalis
B. burgdorferi anti
Enterococcus faecalis
Hydrolytic acid sy
Chlamydia trachoma
Enterococcus faecalis
Enterococcus faecalis
Enterococcus faecalis
Enterococcus faecalis
Enterococcus faecalis
Group B Streptococcus
Enterococcus faecalis
E. coli cellular P
Enterococcus faecalis
Pseudomonas aeruginosa
Salmonella typhi C
Streptococcus pneumoniae
H. pylori GHPD 149
Helicobacter pylori
S. epidermidis ope
Staphylococcus aureus
Putative P. agalactiae
C. glutamicum prote
Putative P. agalactiae
Novel human diageno

DR WPI: 2001-648549/74.
 DR N-PSDB; AAD20238.
 PT Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for
 PT preventing, diagnosing and treating Chlamydia infections in mammals, in
 PT particular humans -
 XX
 XX
 XX Claim 21; Fig 1; 88pp; English.
 CC The present invention relates to novel Chlamydia pneumoniae ATP-binding
 CC cassette protein and its corresponding gene. Sequences of the invention
 CC are useful for detecting Chlamydia infection by assaying a body fluid
 CC of a mammal with the components. They are also used as vaccines. ATP
 CC binding cassette antibodies and vaccines of the invention are useful
 CC for preventing or treating Chlamydia infection e.g. infection caused
 CC by *C. trachomatis*, *C. psittaci*, *C. pneumoniae* or *C. pecorum* in mammals,
 CC such as humans. The nucleic acid molecules are useful for producing
 CC ATP-binding cassettes, in the construction of vaccine vectors such
 CC as poxviruses, which are further useful for preventing and/or treating
 CC Chlamydia infection and in the construction of attenuated Chlamydia
 CC strains that can over-express the nucleic acid molecules or express
 CC it in a non-toxic, mutated form. The present sequence is ATP-binding
 CC cassette from Chlamydia pneumoniae.
 XX
 XX Sequence 532 AA;
 SQ

Query Match 100.0%; Score 2739; DB 22; Length 532;
 Best Local Similarity 100.0%; Pred. No. 2,7e-240;
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISVGCITITLLSVYLOCKKSSHSSTSRGELAINIRDEPRLSDPROVRLSISL 60
 DB 1 mrksvgtictllslsvylgckesshsstsrgealnrdeprslprgvrlsseisl 60
 QY 61 VKHIYEGVQENNTSGNIEPALAEDYSLSDDGLTYFFKLSAFWNGSPDLAEFTSMK 120
 DB 61 vkhiyeglvqennlsgniepalaedyssddgltlyffklsafwngspdlaefteswk 120
 QY 121 QVATQEVSGIYAFALNPTRKTRKIOEGHLSIDHFGVHSPNESTLVLTLESTSHFLKLA 180
 DB 121 qvatqevsgiyafalnptktrkioeghlsidhfgvhspsnestlvltleestshflklla 180
 QY 181 LPVFFPVHKSORTLOSKSLPTASGAFYPRKIKOKIKIRSKPHRYNOSOVETKTTIHF 240
 DB 181 lpfvfpvshsqtllsqskslptasgafypkrikqkwikiskphryngsqveektllthf 240
 QY 241 IPDANTAAFLFNOGKLMNOGPPWGERIPQETLSNLOSKGHLHSPFVAGTSMLEFNINRKP 300
 DB 241 ipdantaaklfnqgklmwgppwgeripqetlsnlgshhlsfivagstswlfnlnktr 300
 QY 301 LNNMKIREALSAIDKREALYSTIFLGRAKTADHLLPTNHSYPRHOKOENAROAYAKKL 360
 DB 301 lnnmkirealasaidekrealvstiflgraktadhllptnhshyprhqgemaqgqayakkl 360
 QY 361 FKRALEELQITAKDLEHNLIFPVSSASSLLVOLLIRQKESIGFAPFVGEFPLLOA 420
 DB 361 fkrealeeqitakdehnlifpvssassllvqllirqkwesigfapfvgefailla 420
 QY 421 DLSNGFSLATGFWFADPMAFLTFAPYPSGVPYALINHKDELFTIQIIEEDOHOKR 480
 DB 421 dlsngfslatgfwfadmfafltfaypsgvpypyalinhkdelftiqiieedohokr 480
 QY 481 SELVSOASILETHIIEPIYHDAPFAMNKLSLWGSPGVVDERYAKEN 532
 DB 481 selvsasilethiiepiyndaigfamnklslnlgspsgvvdfryaken 532

RESULT 2
 AAY34796
 ID AAY34796 standard; Protein: 514 AA.
 XX
 AC AAY34796;

XX 13-SEP-1999 (first entry)
 DE Chlamydia pneumoniae lipoprotein sequence.
 XX
 DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 XX 20-NOV-1998; 98WO-IB01890.
 PE
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GENSET) GENSET.
 XX
 PT Griffiths R;
 XX
 DR WPI: 1999-357842/30.
 XX
 PS Genome sequence of Chlamydia pneumoniae
 PS Page 769-770; Disclosure; 1912pp; English.
 XX
 AAY34584-735879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-735879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 XX
 SQ Sequence 514 AA;

Query Match 96.6%; Score 2645; DB 20; Length 514;
 Best Local Similarity 100.0%; Pred. No. 9e-232;
 Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISVGCITITLLSVYLOCKKSSHSSTSRGELAINIRDEPRLSDPROVRLSISL 60
 DB 1 mrksvgtictllslsvylgckesshsstsrgealnrdeprslprgvrlsseisl 60
 QY 61 VKHIYEGVQENNTSGNIEPALAEDYSLSDDGLTYFFKLSAFWNGSPDLAEFTSMK 120
 DB 61 vkhiyeglvqennlsgniepalaedyssddgltlyffklsafwngspdlaefteswk 120
 QY 121 QVATQEVSGIYAFALNPTRKTRKIOEGHLSIDHFGVHSPNESTLVLTLESTSHFLKLA 180
 DB 121 qvatqevsgiyafalnptktrkioeghlsidhfgvhspsnestlvltleestshflklla 180
 QY 181 LPVFFPVHKSORTLOSKSLPTASGAFYPRKIKOKIKIRSKPHRYNOSOVETKTTIHF 240
 DB 181 lpfvfpvshsqtllsqskslptasgafypkrikqkwikiskphryngsqveektllthf 240
 QY 241 IPDANTAAFLFNOGKLMNOGPPWGERIPQETLSNLOSKGHLHSPFVAGTSMLEFNINRKP 300
 DB 241 ipdantaaklfnqgklmwgppwgeripqetlsnlgshhlsfivagstswlfnlnktr 300
 QY 301 LNNMKIREALSAIDKREALYSTIFLGRAKTADHLLPTNHSYPRHOKOENAROAYAKKL 360
 DB 301 lnnmkirealasaidekrealvstiflgraktadhllptnhshyprhqgemaqgqayakkl 360

	Query Match	41.18;	Score 1126;	DB 20;	Length 494;	
	Best Local Similarity	43.8%;	Pred. No.	1.se-93;		
	Matches 217;	Conservative 104;	Mismatches 168;	Indels 6;	Gaps 3	
QY	41 RDEPNSLPDROYRLSEISLVKHIYEGLVQENNLSPNEPALAEADYSASSDSOTITYPEIK 100	:	::	::::: :	:	::
Dd	1 kempisrlipgkrtlladqtlmrhlmyglvveehsqngsklpalesesrlyleagdrtytylefk 60	:	:	: ::: :	:	::
QY	101 SAEWSGSDPLAAEDFISSWKQVAIVDVEGVCIATFLNLIKVKRIQLQGCHLSIDFGVVSPNP 160	:	:	:	:	:

[illegible]

Query Match	24.8%;	Score 680.5;	DB 20;	Length 529;
Best Local Similarity	34.0%;	Pred. NO. 5.8e-53;		
Matches 179;	Conservative 92;	Mismatches 226;	Indels 29;	Gaps 10

[illegible]

RESULT 7
22X07040

ID	AA	97040	standard; Protein; 544 AA.
XX			

AC AAY97040;

DT	31-OCT-2000	(first entry)
XX		

B. subtilis oppA ligand binding protein.

nm Upp operon; Spook; oligopeptide permease; sporulation; ABC transporter;
KW ATP-binding cassette transporter; putative

opprb; opprc; opprd; opprf; ligand binding protein..
xx

XX
SUBLIS

XX
XX

XX

XX	2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
----	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

XX
XX
DA (CENT) / CONTINUED FROM PREVIOUS PAGE

PA (DIAZ/) DIAZ-TORRES M.
PA (FERRE/) FERREIRA D

XX Diaz-Torres M Ferrari F

XX
DR
WPT: 2000-452412/39

XX N:PSDB; AAA51804.

cytokines, in gram-p

genes of the opp operon gene cluster

XX
XX

The *opp* operon of *Bacillus* (also known as *spoK* operon) encodes an oligopeptide permease that is required for the initiation of sporulation and the development of genetic competence. The *opp* operon is a member of the family of *Ara*-binding cassette (ABC) transporters involved in the import or export of oligopeptides from 3-5 amino acids. *Bacillus* strains containing a mutation in the *opp* operon produce more recombinant protein than the wild-type strain. Therefore, gram-positive microorganisms, especially *Bacillus* strains, containing a mutation in at least one of the genes of the *opp* operon can be used for heterologous protein production, especially hormones, enzymes (preferably proteases such as subtilisin), growth factors or cytokines.

Sequence 544 AA;

Query Match 22.5%; Score 617.5; DB 21; Length 544;
Best Local Similarity 29.4%; Pred. No. 3.3e-47;
Matches 159; Conservative 107; Mismatches 249; Indels 25; Gaps 13

```

10 ITILSLSVLÖGESSHSSSRGE-----LAINRDEPRSLDPROVRLISEISLV 61
:::| | | | | :
:
```

[illegible][illegible][illegible]

127 waldnessey wikkasat-1-133

QY 181 LPFPPVHKSORTLQSKSLPIASGAFYPRN--IKOKOM-----IKISKPHHYNSQVET 233
 Db 187 fytympink kiaeakmkwmtlnagdyvsngpfkmtawkhsgstltkxndgywkdvkyl 245
 QY 234 KTTTHPTPANTAAKLFNOGKLNOCPPMGERIPQETLSNLSKGLHSFVACTSWLT 293
 Db 246 kldhwmnnnnnltkfkagagdaagmldqg-lptestptlkdgslhvepiagvywk 304
 QY 294 FNIKKFPLNNMKLREALSALDKEALVSTIFLGRAKTADHLPTNHSYEPKOEKQEMAKOR 353
 Db 305 fnteaqldhvnrlkaltyldrgskivknvqgqjpaamaaavppcmkfgedhkeyfkdn 364
 QY 354 QA-VAKKLFEALFEELQIN-AKODLHNLIPVSSASSLVLQUREQKESLGPALPIV 411
 Db 365 dvktakeylekylkenglskaskdipkiklsyn-tddahakiaqavgemkknlgdveld 423
 QY 412 GKFEALLQADUSGNSFLATGWFADPAPMAFLTIFAVPS-GVPPYAINKDELLEON 470
 Db 424 nsewnyidklhngdyqgrmgwlgdndpfnfelfrckngndtqwenpefkllng 483
 QY 471 IEQBOHOKRSELYSQA-SYLEFPHLEPIYHDAPQFAMNKLISNLGVSPTGVDFRYA 529
 Db 484 sqetcdktknaellkkaegltidemp-vabilyfyldtwgdenlkvymptgvyfyna 542

RESULT 8
 AAY37004
 ID AAY37004 standard; Protein: 214 AA.
 AC AAY37004;
 DT 07-OCT-1999 (first entry)

Chlamydia trachomatis transport protein.
 Vaccine: eye disease; conventional trachoma; nonendemic trachoma;
 paratrachoma; inclusion conjunctivitis; genital disease; perhepatitis;
 nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

Chlamydia trachomatis.
 OS
 PN WO9928475-A2.
 PD 10-JUN-1999.
 PF 27-NOV-1998; 98WO-IB01939.
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97ER-0015041.
 PR 17-DEC-1997; 97ER-0016034.

PA (GEST) GENSET.
 XX
 PI Giffais R;
 XX
 DR WPI: 1999-371125/31.
 PT
 PS Genome sequence of Chlamydia trachomatis
 XX
 PS Disclosure; Page 836; 1755pp; English.

AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 of Chlamydia trachomatis (see A4201425). The polypeptides can be used as
 vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 can also be used to control growth of the microorganism. Chlamydia
 trachomatis is responsible for a large number of diseases, e.g. eye
 diseases such as conventional trachoma, nonendemic trachoma,
 paratrachoma, and inclusion conjunctivitis; genital diseases such as
 nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perhepatitis, Bartholinitis, pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.
 XX
 SQ Sequence 214 AA;

Query Match 19.0%; Score 521.5; DB 20; Length 214;
 Best Local Similarity 48.5%; Pred No. 4.2e-39;
 Matches 98; Conservative 41; Mismatches 62; Indels 1; Gaps 1;

QY 331 ADHLPTNHSYEPHOKQENARQAVAKKLFEALFEELQINAKDELHNLIPVSSASS 390
 Db 14 akhlipahhhyepqpskyggaatlaksllealtelmtkileklyplftsactvns 73
 QY 391 LVVOLIREQKESLGPALPIVGEKFEALLQADUSGNSFLATGWFADPAPMAFLTIFAY 450
 Db 74 qiaqmlrqwrrslgltfplcsgkeyallqndlgntfmslgygfadfsdplafstfs- 132
 QY 451 PSQVPPYAINKDELLEONIEQBOHOKRSELYSQA-SYLEFPHLEPIYHDAPQFAMN 510
 Db 133 skgvkpyalqgpdqdlisleteknpkrtsaliseaslyterqvnelyndvthytin 192
 QY 511 KKLSTLGVSPGVDFRYAKEN 532
 Db 193 nklstfvrilpssglvdmryaks 214

RESULT 9
 AAY20067
 ID AAY20067 standard; Protein: 512 AA.
 AC AAY20067;
 DT 19-JUL-1999 (first entry)

B. burgdorferi antigenic protein, 1606.aa.
 DE
 XX
 KW Antigenic protein; vaccine; Lyme disease; infection; detection.
 KM
 XX
 OS
 PN WO9859071-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; 98WO-US12718.
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI (MEDI-) MEDIMUNE INC.
 XX
 PI Choi GH, Erwin AL, Hanson MS, Iathigra R;
 XX
 DR WPI: 1999-189980/16.
 DR N-PSDB; AAX61764.
 XX

New isolated Borrelia burgdorferi nucleic acids - used to develop
 products for the diagnosis, prevention and treatment of diseases
 caused by Borrelia, particularly Lyme disease

Claim 12; Page 182; 275pp; English.

This sequence represents a Borrelia burgdorferi (Bb) protein of the
 invention, which is suitable for use in a vaccine. The Bb polypeptides
 can be used in vaccines for eliciting protective antibodies to members of
 the Borrelia genus, particularly for the use against Lyme disease in
 humans and animals. They can be used for preventing or attenuating an
 infection caused by a member of the Borrelia genus. The products can also
 be used for detection of members of the Borrelia genus.

Age Group	2006 (%)	2008 (%)
18-29	85	80
30-49	80	75
50-69	75	70
70+	70	65

KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KM vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

XX WO9227105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

XX Griffiths R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 771-772; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
XX frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
XX C. pneumoniae causes respiratory disease such as pneumonia and
XX bronchitis and is thought to be a contributing factor in heart
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX nodosum or pharyngitis. The polypeptides encoded by the open reading
XX frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX nucleotide sequences can also be used as immunogenic compositions,
XX especially where the vector directs the expression of a neutralising
XX epitope of C. pneumoniae.

XX Sequence 448 AA;

Query Match 17.6%; Score 482.5; DB 20; Length 448;
Best Local Similarity 27.0%; Pred. No. 4,7e-35;

Matches 141; Conservative 92; Mismatches 190; Indels 99; Gaps 17;

QY 10 ITLLSLSVLQGCESHSSTSGELAINIDPRSLDPROVRLSEISLVKHYELV 69

DB 19 ILLFL-ILSLTGC--SYSSKHKQSLIPIHNDPVALISPEQAKRAMDISIAGLIIFDGLT 75

QY 70 QENNTLSGN-IEPALAEDYLSLSDGLTYTFKLK-SAFWSNGDPPLAEDFESKKQVATQEV 127

DB 76 rethresndlelaiaasyvedfcsyffikdsalvsdgtpltsedlnave----- 128

QY 128 SGIVAFALNPKNVKKIQEGHLSI-DHGVHSPWESLVVLTLESPTSHPKLKLALPVFP 186

DB 129 --yagensp-----hqlfnglnfstpsnatlthdspdpdkllarpa-fa 174

QY 187 VHSQRTLOSRLPIASGAFYPKNTKOKWKIKLSKNHYNQSQVETKITTITHTPRDANT 246

DB 175 ifkxepn-kifsgpyrlvefyghn-----ihkxmpnydydhwvnsinsixk111ipdyt 228

QY 247 AAKLFNGKLNMGPPWGERIPQETLSNLSQKGLHSFDVAGISMLTNINKPPLNNMKL 306

DB 229 aihlimgkvdwgqpwbgdipwel--hkqgyhytylvevgaftclntnspnldnq 286

QY 307 REAASALDEALVSTIFIGRAKTRADHLPLNINHSYBEHQKQDMAQRQAVAKKLFKDALE 366

DB 287 rhratcidrslieea1ggtqpaetl-----srgapqpnqyvk----- 326

QY 367 ELQTTANDEHLNLTIFPVSSASSSLVOLTIREQWKESLFAPIYIGKFFALQADLSSGN 426

DB 327 --qkpltpqekivltypsdillrqriaelikeqk-aagidillleglyhlfvfkxkvqd 383

QY 427 FSLATGGMFRADRADMAFLITFAVPSSGVPPYAIINHKDFLEIIQNTIEQDQHQKSELVSG 486

DB 384 yaia-----qtgyayypg-----anlise 403

QY 487 ASLVETPHIIEPIYHDARQFAMNKKLSNLGVSPTGVVDERY 528

DB 404 edkllgmfeil-plyllysydytlgtqfiegvylnyngsvadlky 444

RESULT 12

AAI00030

AAI00030 standard; Protein; 547 AA.

AAI00030;

20-APR-1999 (first entry)

Enterococcus faecalis protein EF012.

Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

detection; attenuation; antigenic.

Enterococcus faecalis.

WO9850554-A2.

12-NOV-1998.

04-MAY-1998; 98WO-US08959.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046555.

(HUMA-) HUMAN GENOME SCI INC.

Bailey C, Choi GH, Hromockyj A, Kunsch CA;

WPI; 1999-070095/06.

N-PDB: AAX20020.

New isolated Enterococcus faecalis polynucleotides - used to develop

products for the detection of Enterococcus and for use in vaccines

for prevention or attenuation of Enterococcus infection

Claim 9; Page 90, 301pp; English.

The present sequence represents a protein isolated from

Enterococcus faecalis. The present invention describes genes, proteins

and antigenic polypeptides isolated from E. faecalis. The proteins can

be used in vaccines for preventing or attenuating an infection caused

by a member of the Enterococcus genus in an animal. They can also be

used for detecting Enterococcus antibodies in a sample. The nucleotide

sequences can be used for detecting Enterococcus nucleic acids.

Products from the present invention can also be used for screening

compounds to identify agonists and antagonists of E. faecalis protein

activity.

Sequence 547 AA;

Query Match 16.6%; Score 455; DB 20; Length 547;
Best Local Similarity 25.2%; Pred. No. 2e-32;

Matches 136; Conservative 111; Mismatches 256; Indels 36; Gaps 15;

QY 4 ISVGICITILSLSVLQGC-----KESHSSTSGELAINIDPRSLDPROVRL 55

DB 1 mklgkvvgylatgfliaacgylkaeaekvdsnglaaeqslsspspistldttqttdk 60

QY 56 SEISLVKHYEGVVOENNSGNIEPALAEDYLSLSDGLTYTFKLKAF-WSNGDPLAED 114

DB 61 ntfcmeghifeglyrtdodsacv-palakdvksiddgrkyhftlreglkwsngpdltagd 119

QY 115 FIESMKQVATQEVSGYIAFALNPINVKRIQEGHLSIDHGVHSPWESLVVLTLESPTS 174

Db 120 fVyswKklvlpatlgpnaylldsvknstfeirngeksvdelglsapndkefiwelkqags 179
 QY 175 FLKILALPVPFVPHKSORLQSKSLP-----IASGAFYPKN--IKOKWIKLSKNPHY 226
 Db 180 flavvsLawlapnqkfveaqqkdyalalsehllyspgrflaanwdatsdtw-tilkknpeyy 238
 QY 227 NOSOVERKTIITFIPIIDANTAAKLFNGLKMWGCPFWGGERIPQETLSNLQ--SKGHLHSFD 265
 Db 239 dadqvkleeavastikehtgtinlyvneidl-----vrlngqyvgqddppgyvshpd 292
 QY 286 VAGTSWLTFINNK-FPLNNMKLREALASALDKEALVSTIFLGRATADHLLPTNHSYPE 344
 Db 293 va-nyflidfnkkkeqrpilanvrlkaigaldkeelcgsvlndgsprlmglllpsklyanpe 351
 QY 345 HOKEMAQRARAKKLFKALEELQITAKDL-EHLNL-IFPVSSASSLLVQLIREQWKE 402
 Db 352 tdeeftrayageylikndvkkagaewtkaqadvgrkvlslaaadtqgkxiieyqsglqge 411
 QY 403 SL-GFAIPYIGKEFALLQDLSGNSFATGWFADFPAPMAFLTIFAPGVPYAINH 461
 Db 412 nlpgleltlssqpsnmvngstreknyelstsgwlaqsseldsyinlygeassy-ngynyh 470
 QY 452 K---DFLEILQNIQSDHOKRSELVSOASLYLETFHIEPIYHDAROFAMNKKLSNL 516
 Db 471 nakydqiveearltinampekqfaeykkaedlllngdaaqvpylygsasnylinpkikgi 529

RESULT 13

AAV00031 standard; Protein; 528 AA.

AAV00031;
 20-APR-1999 (first entry)
 Enterococcus faecalis antigenic polypeptide fragment EF012.

Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic.

Enterococcus faecalis.

WO9850554-A2.

12-NOV-1998.

04-MAY-1998; 98WO-US08959.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

Bailey C, Choi GH, Hromockyj A, Kunsch CA;

WPI; 1999-070095/06.

N-PSDB; AAX20021.

New isolated Enterococcus faecalis polynucleotides - used to develop
 products for the detection of Enterococcus and for use in vaccines
 for prevention or attenuation of Enterococcus infection

Claim 9; Page 91; 301pp; English.

The present sequence represents an antigenic polypeptide fragment
 isolated from Enterococcus faecalis. The present invention describes
 genes, proteins and antigenic polypeptides isolated from E. faecalis.
 The proteins can be used in vaccines for preventing or attenuating an
 infection caused by a member of the Enterococcus genus in an animal.
 They can also be used for detecting Enterococcus antibodies in a sample.
 The nucleotide sequences can be used for detecting Enterococcus nucleic
 acids. Products from the present invention can also be used for

CC screening compounds to identify agonists and antagonists of E. faecalis
 CC protein activity.
 XX

Sequence 528 AA;

Query Match 16.5%; Score 452.5; DB 20; Length 528;
 Best local similarity 26.6%; Pred. No. 3,36-32;
 Matches 138; Conservative 102; Mismatches 244; Indels 35; Gaps 17;

QY 22 GKESHSSTSRGELA---INIRDEP--RSLDPQVLLSEISLVHYEGVLENNLS 75
 Db 3 glkeaaekvds-gnlaaeqkisspsapstldtqtckntfmeqhlfeiglyrtddds 61
 QY 76 GNIEPALAEYSLSDDGLTYTFKLKSAF-WSNGDEPLAEDFISWQVATOEVSQYVFA 134
 Db 62 atv-palakdkvksiddgkfhflfreglkwngpeltagdtlyvswkklvtptatlgpnayl 120
 QY 135 LNPINKVAKIEGHLSTDEHGVHSPNESTLVVTLSEFSLKLLALPVFFVPHKSORTL 194
 Db 121 ldsyknstfeirngeksvdelgisapndkefiwelkqagsflavvsLawlapnqkfvea 180
 QY 195 QSKSLP-----IASGAFYPKN--IKOKWIKLSKNPHYNOGQVETKTIHFIPDANT 246
 Db 181 qkdyaidselhllysgpfilanwdatsdtw-tilkknpeyydaaqvaveavstikeht 239
 QY 247 AAKLFNOKKLNWGCPFWGGERIPQETLSNLQ--SKGHLHSFVAGTSMALFINNK-FPLNNM 304
 Db 240 glnlyvneidl-----vrlngqyvgqddppgyvshpdva-nyflidfnkkkeqrpilanv 292
 QY 305 KLREALASALDKEALVSTIFLGRATADHLLPTNHSYFEHOKEMAQRARAKKLFKEA 364
 Db 293 hlrtkaigaldkealqsvlndgsprlmglllpsklyanpedtrayageylikndvkk 352
 QY 365 LLELQITAKDL-EHLNL-IFPVSSASSLLVQLIREQWKE-LGFAIPYIGKEFALLQAD 421
 Db 353 qawetkaqadvgrkvlslaaadtqgkxiieyqsglqgenlpgleltlssqpsnmng 412
 QY 422 LSSGNSFATGWFADFPAPMAFLTIFAPGVPYAINHK---DFLEILQNIQSDH 477
 Db 413 treknyselstsgwlaqsseldsyinlygeassy-ngynyhakydqiveearltinampe 471
 QY 478 OKRSELVSOASLYLETFHIEPIYHDAROFAMNKKLSNL 516
 Db 472 kqfaeykkaedlllngdaaqvpylygsasnylinpkikgi 510

RESULT 14

AAV19870 standard; Protein; 523 AA.

AAV19870;

19-JUL-1999 (first entry)

B. burgdorferi antigenic protein, f607.aa.

Antigenic protein; vaccine; Lyme disease; infection; detection.

Borrelia burgdorferi.

WO9859071-A1.

30-DEC-1998.

18-JUN-1998; 98WO-US12718.

03-SEP-1997; 97US-0057483.

20-JUN-1997; 97US-0050359.

22-JUL-1997; 97US-0053344.

22-JUL-1997; 97US-0053377.

(HUMA-) HUMAN GENOME SCI INC.

PA (MEDI-) MEDIMUNE INC.

XX Chai GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI: 1999-189980/16.

DR N-PSDB: AAX61567.

XX

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 products for the diagnosis, prevention and treatment of diseases
 caused by *Borrelia*, particularly Lyme disease

PS Claim 12; Page 97-98; 275pp; English.

CC This sequence represents a *Borrelia burgdorferi* (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the *Borrelia* genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the *Borrelia* genus. The products can also
 CC be used for detection of members of the *Borrelia* genus.

XX Sequence 523 AA;

Query Match 15.5%; Score 425.5; DB 20; Length 523;

Best Local Similarity 23.4%; Pred. No. 9.2e-30;

Matches 122; Conservative 131; Mismatches 217; Indels 51; Gaps 15;

QY 1 MRKISVGICITLLSLVVLGGCKSSSTSRGELAINIRDEPRSDPRVRLLESLI 60

DB 1 mkykialmlifslaci-----snakektivrvsnlsepssldpqlstldlygsnl 53

QY 61 VKHIEGLVQENNLSGNTPEALADYSLSDGILTYFKLK-SAEWSNDPLTAEDETESM 119

DB 54 Ithflglavkdsqgkykpglakswnisedgilyfnlredlwsdgvateelkky 113

QY 120 KOVATQGVGIAFAL-NPIKNVAKIOEGHLSIDHGVSNPNESTIVVTIESPTSLFKL 178

DB 114 Irlkktaamyanaikstiknaeyfdetvpeselgikaidsktleitlspkpyfpdm 173

QY 179 LAIPVFPV--HRSQRTLOSKSLP---IASGAFYPRNKIKOKOWIKLSKNPHYVNOQVET 233

DB 174 Ithsayipymlivekygemwtnpenlivsgaykikersindkivieknyaknavei 233

QY 234 KTTIHFTIDANTAKLFPNOGKLNW-OGPPWGERIPOETLSNLOSKGHLSPFVAGTSL 292

DB 234 dev-ityptegsvaymyingeldfig-----aeknlleeikliddysgikngmayi 286

QY 293 TFINKFPLNMKTLREALASALDKREALVSTIFLGRAKTADHLPT-NHISY-----P 343

DB 287 afittikpdlmkvrgaisaiaidrecltkvvlkgsdptnltpkldysgknl1ltdp 346

QY 344 EHOQDMARQAVAKLFEKALBELQITAKDLEHNLTFPVSASASLVLQILREQMES 403

DB 347 enakkllaeagypdgkfp-----tlkykiseqrpt-taeftigeqfki 389

QY 404 LGFALPIYGRFALLQDLSSGFLSLATGCFADPAPMAFL-TIPAYPSG-VPPYALNH 461

DB 390 Ithlelenewtftfgsrrtgnysvgyidgfdplfildslfthenhrlgkyksn 449

QY 462 KDFLEILNINQEQDHQKRELYSOA-SLYETFHIIPEIY 501

DB 450 keydalikknsfelpdikrgdillrqeeleakdfmaply 490

RESULT 15
 AAY19871
 ID AAY19871 standard; Protein; 506 AA.

XX AAY19871;

AC AAY19871;

DT 19-JUL-1999 (first entry)

XX

DE B. burgdorferi antigenic protein, t607 aa.

XX Antigenic protein; vaccine; Lyme disease; infection; detection.

XX *Borrelia burgdorferi*.

XX WO9859071-A1.

XX 30-DEC-1998.

XX 18-JUN-1998; 98MO-US12718.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (MEDI-) MEDIMUNE INC.

XX Chai GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI: 1999-189980/16.

XX N-PSDB: AAX61568.

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 products for the diagnosis, prevention and treatment of diseases
 caused by *Borrelia*, particularly Lyme disease

PS Claim 12; Page 98; 275pp; English.

CC This sequence represents a *Borrelia burgdorferi* (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the *Borrelia* genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the *Borrelia* genus. The products can also
 CC be used for detection of members of the *Borrelia* genus.

XX Sequence 506 AA;

Query Match 15.4%; Score 421; DB 20; Length 506;

Best Local Similarity 24.6%; Pred. No. 2.2e-29;

Matches 118; Conservative 119; Mismatches 198; Indels 44; Gaps 14;

QY 43 EPSLDPQVRLRLSEISLVKHYEGLVQENNLSGNTPEALADYSLSDGILTYFKLK-S 101

DB 19 epsltdpqlstldlygsnlitnlflglavkdsqgkykpglakswnisedgilyfnlred 78

QY 102 AFSNNDPLTADFTESKQVATQVSGIYAFAL-NPIKNVAKIOEGHLSIDHGVHSPN 160

DB 79 Iwsgdvaitaeelkysylrlnktaamyanaikstiknaeyfdetvpeselgikaid 138

QY 161 ESTLVLTESPTSHFLKLLALPVPFPV--HRSQRTLOSKSLP---IASGAFYPRNKIKOKO 215

DB 139 sklletltsppryfpdmIthsayipymlivekygemwtnpenlivsgaykikersind 198

QY 216 WIKLSKNPHYVNOQVETKTTIHFTIDANTAKLFPNOGKLNW-OGPPWGERIPOETLSN 274

DB 199 Ithlelenewtftfgsrrtgnysvgyidgfdplfildslfthenhrlgkyksn 311

QY 275 LQSGHLSFVAGTSMWLTJFNINKFPLNMKTLREALASALDKREALVSTIFLGRAKTADHL 334

DB 252 Iklrddyysgikngmayiafntlkpdlmkvrgaisaiaidrecltkvvlkgsdptnl 311

QY 335 LPT-NHISY-----PEHOKQDMARQAVAKLFEKALBELQITAKDLEHNLTFPVS 385

DB 312 tpfkfdysgknl1lfdpenakllaeagypdgkfp-----tlkykis 355

QY 386 SSASSLVLQILREQMESLGFAPIYGRFALLQDLSSGFLSLATGCFADPAPMAFL 445

DB 356 egprt-taeftigeqfkiIthlelenewtftfgsrrtgnysvgyidgfdplfcl 414

OY 446 -TIFAYPSG-VPEYAINHKDFLEITQNEQODHQRSELSQA-SLYLETFHIIETPY 501
Db 415 dslfthenhlgaykysnkeydaliksnfelcblpikrgdillrgaeelisekdffmapply 473

Search completed: July 26, 2002, 04:38:09
Job time: 5787 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 03:34:32 ; Search time 32.07 Seconds
(Without alignments)
405,189 Million cell updates/sec

Title: US-09-824-567-2
Perfect score: 2739
Sequence: 1 MRRISVIGITITLLSLSVTL.....LSNLGVSPGVDFRYAKEN 532

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCRNUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match length	DB ID	Description
1	280.5	10.2	642	US-08-245-511-48
2	280.5	10.2	642	US-08-600-993A-48
3	270.5	9.9	643	US-08-245-511-47
4	270.5	9.9	643	US-08-600-993A-47
5	165.5	6.0	556	US-09-385-028-9
6	108	3.9	398	US-08-935-263-12
7	107	3.9	1089	US-08-375-300-2
8	107	3.9	1089	US-09-177-431-2
9	107	3.9	1089	US-08-328-256-10
10	104.5	3.8	557	US-08-471-454-2
11	104.5	3.8	557	US-08-466-974-2
13	104.5	3.8	557	US-08-471-453-2
14	104.5	3.8	557	US-08-307-588-4
15	104.5	3.8	2071	US-09-415-522-6
16	103	3.8	436	US-08-660-347-2
17	102	3.7	2237	US-08-354-973-1
18	100.5	3.7	580	US-08-188-930-307
19	100	3.7	1257	US-08-049-783-2
20	100	3.7	1257	US-08-158-732-6
21	100	3.7	1257	US-08-304-626-6
22	100	3.7	1257	US-08-316-301A-6
23	100	3.7	1257	US-08-611-928-6
24	100	3.7	1257	US-09-173-891-6
25	100	3.7	1257	US-09-076-137-6
26	100	3.7	1257	PCT-US92-03624-6
27	99	3.6	476	US-08-378-313-34

28	97	3.5	1664	US-08-642-846-2	Sequence 2, Appl1
29	97	3.5	1664	US-09-264-604-2	Sequence 2, Appl1
30	97	3.5	2710	US-08-480-604A-6	Sequence 6, Appl1
31	97	3.5	2710	US-08-405-496A-6	Sequence 6, Appl1
32	97	3.5	2710	US-08-915-136-6	Sequence 6, Appl1
33	94	3.4	1142	US-08-993-118-7	Sequence 7, Appl1
34	94	3.4	1142	US-08-845-828C-7	Sequence 7, Appl1
35	94	3.4	1142	US-09-061-709-2	Sequence 2, Appl1
36	93.5	3.4	2629	US-08-751-189-4	Sequence 4, Appl1
37	93.5	3.4	2629	US-09-060-836-4	Sequence 4, Appl1
38	93.5	3.4	2629	US-09-184-445-4	Sequence 4, Appl1
39	92.5	3.4	893	US-07-977-434-4	Sequence 4, Appl1
40	92.5	3.4	893	US-08-458-819-4	Sequence 4, Appl1
41	92.5	3.4	893	US-09-105-697-10	Sequence 4, Appl1
42	92.5	3.4	893	PCT-US91-07035-4	Sequence 10, Appl1
43	92	3.4	512	US-08-856-253-6	Sequence 6, Appl1
44	91.5	3.3	920	US-08-930-996A-8	Sequence 8, Appl1
45	90	3.3	776	US-08-961-083-86	Sequence 86, Appl1

ALIGNMENTS

RESULT 1
US-08-245-511-48
Sequence 48, Application US/08245511
Patent No. 5928900
GENERAL INFORMATION:
APPLICANT: Masure, H Robert
APPLICANT: Pearce, Barbara J
APPLICANT: Tuomenen, Elaine
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
TITLE OF INVENTION: CELLULAR VACCINES BASED THEREON
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 642 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
IMMEDIATE SOURCE:
CLONE: amla

```

FEATURE:
OTHER INFORMATION: NOTE: the reference contains a
OTHER INFORMATION: sequence error; the correct sequence shown below is obtained
OTHER INFORMATION: from GENBANK
PUBLICATION INFORMATION:
AUTHORS: Allouing, et al.
JOURNAL: Mol. Microbiol.
VOLUME: 4
PAGES: 633-644
DATE: 1990
US-08-245-511-48

Query Match
Best Local Similarity 10.2%; Score 280.5; DB 2; Length 642;
Matches 137; Conservative 104; Mismatches 243; Indels 147; Gaps 26;

QY 19 VLQCKESHSTSGELAIN--IRDEPRSLDPQVRLISELSVYKHIEGLVQENNLG 76
D 2 VLAC--SSKSSSDSAPKAGVYTTADPETIDYLSRKNSYVTSNGIDGLFTNDY-G 59
QY 77 NIEPALADYSLSDGLTYTFKLSA--FWSNGD--PLTAEDFIESMKOVATQEVSGI 130
D 60 NLAPVAEDMEVSKDGLTYTKIRKGVKMTSDGEVAVTAADPVNGIKHAADKSEAM 119
QY 131 YAFALNPKNVRIQEG-HLSIDHGVHSPNESTLVLTLESPTSHFLKLLAPVFPYHK 189
D 120 Y-LAENSVKGLADYLSGTSTDFSVGVKAVDDYTLQTLNQPDPFNWSKLTYSIFWPLNE 178
QY 190 SORTLOSKSLP-----IASGAFYPRNIKOKWIKLSKNPHYVNSQVETKTTT--- 238
D 179 EETSCKSDPAKPTDPTSLYXNPFILKGLTRKSSVEPVKNQYWDKEVHLDITNLAYY 238
QY 239 -----HETPDANTAKLFENQKLNMOGPPWGRIPQETLSMLQSKGHLHSDVAG 288
D 239 DGSQDESLEKFTSGAVARLYPTSS-NY-----SKVAEYKMDI-----YTIQSGSG 286
QY 289 TSMLETFNKNKPLN-----NKKLEALASLDKEA----- 318
D 287 IAGLGVNINDQSTNYTSKTTTDESEKVAATKALLNDFRQALNFDLSRAYSQAQINGKOGAA 346
QY 319 -LVSTIFL-----GAKTADHLPTLNHSYPEHQ-----QEAQAQAATAKLPREA 364
D 347 LAVRNLFPKDPFVASEKTEFGDLVAAQPAVGDEMKRVNLAQODQLFNADKAAEFKKA 406
QY 365 LEELOITAKMLEHLNLFYVSSASSLL--VOLIREQWKSJGFALPIV-----GKEF- 415
D 407 KKALEADGVOP-THLDVAVDASKNYISRIQSFQSEVETVGVENVVVDIQWMTSDEFL 465
QY 416 -ALLQADLSGNSFLATG-GWFAFDAPMAFLFI-----PAYPGVPPYA 458
D 466 NITYYANASSEDMDVSGSVGSPYQDPSTYLDILKTYSSETTYTGAFGNPSPSVYQ 525
QY 459 INHNDPELLONTDEP-ODHOKRESEVSOASLYLETFHIEPIYHDAQFPMNKKSLNLG 517
D 526 VGLEKVDKLDVDAKETSDFNVRYEKYAAQAAMLTDSLSLFI-----AMASSG 573
QY 518 VSP-----TGCV-----YDPRYAK 530
D 574 AAPVLSRYVPTFGASQGTGSKGSDVYFKYK 604

RESULT 2
US-08-600-993A-48
Sequence 48; Application US/08600993A
Patent No. 5981229
GENERAL INFORMATION:
APPLICANT: Masure, H Robert
APPLICANT: Pearce, Barbara J
APPLICANT: Tuomanen, Elaine
TITLE OF INVENTION: BACTERIAL EXPORDED PROTEINS AND
CELLULAR VACCINES BASED THEREON
NUMBER OF SEQUENCES: 59

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,993A
FILING DATE: 1-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 US
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEO ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 642 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOHECTICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
IMMEDIATE SOURCE:
CLONE: amla
FEATURE:
OTHER INFORMATION: the reference contains a sequence error; the
PUBLICATION INFORMATION:
AUTHORS: Allouing, et al.
JOURNAL: Mol. Microbiol.
VOLUME: 4
PAGES: 633-644
DATE: 1990
US-08-600-993A-48

Query Match
Best Local Similarity 10.2%; Score 280.5; DB 2; Length 642;
Matches 137; Conservative 104; Mismatches 243; Indels 147; Gaps 26;

QY 19 VLQCKESHSTSGELAIN--IRDEPRSLDPQVRLISELSVYKHIEGLVQENNLG 76
D 2 VLAC--SSKSSSDSAPKAGVYTTADPETIDYLSRKNSYVTSNGIDGLFTNDY-G 59
QY 77 NIEPALADYSLSDGLTYTFKLSA--FWSNGD--PLTAEDFIESMKOVATQEVSGI 130
D 60 NLAPVAEDMEVSKDGLTYTKIRKGVKMTSDGEVAVTAADPVNGIKHAADKSEAM 119
QY 131 YAFALNPKNVRIQEG-HLSIDHGVHSPNESTLVLTLESPTSHFLKLLAPVFPYHK 189
D 120 Y-LAENSVKGLADYLSGTSTDFSVGVKAVDDYTLQTLNQPDPFNWSKLTYSIFWPLNE 178
QY 190 SORTLOSKSLP-----IASGAFYPRNIKOKWIKLSKNPHYVNSQVETKTTT--- 238

```



```

Db 179 EFETSKSGDFAKPTDPTSLLYNGPFLKGLTAKSSVEFEKNGQYMDKENVHDTINLAY 238
QY 239 -----HETPDANTAKLFPNGKLNMOGPPWGERITQETLSLQSGHLSFDVAG 288
Db 239 DCSQDESIERNFTSGAYSTARLYPTSS-NY-----SKVAEEFKDNI-----YYQSGSG 286
QY 289 TSWLTFNINFPPLN-----NMKLEALASALDKEA-----318
Db 287 IAGGVNIDRQSYNTSKTTDEKATKALKLNKDFQALNFPALDKRSAYSQINGKDGA 346
QY 319 -LVSTIYL-----GKATADHLPTNHSYPEROK-----QEMARQAVAKKLPKEA 364
Db 347 LAVNLFLEVKEDFVSAGEKTFGLVYAAQAPAYGDEMKGYNLADQOGGILNNAKKAEPKKA 406
QY 365 LEELQITAKCLEHLNLEFPVSSASLSL--VQLIREQKESIGPAIPY-----GKEF- 415
Db 407 KKALEADGVQFP-IHLDPVVDQASKNYISRSQSFQSVETVLGVENVVVDIQQMSDFL 465
QY 416 --ALLQADSSGNSFIATNG-GMFADFPADPMAPLTI-----FAYPSGVPPYA 458
Db 466 NITYAANASSEDMDVSGVSGMGPDYQDPSTYIDLKTSSTETTYTGFDNPNPSVVO 525
QY 459 INHDELELLQNIQEQ-ODHOKRSELYSQASLYLEFFHIEPIYHDAQFAMNKKLSNLG 517
Db 526 VGLKEIDKLVDENAKETSDFNVRKYAAQAWLTDSSLIFP-----AASSG 573
QY 518 VSP-----TGV-----VDFRYAK 530
Db 574 AAPVLSRIVPTGASAGTSGSGSNVYTKYK 604

RESULT 3
US-08-245-511-47
; Sequence 47, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; type: amino acid

```

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-245-511-47

Query Match 9.9%; Score 270.5; DB 2; Length 643;
Best Local Similarity 21.6%; Pred. No. 5,4e-17;
Matches 130; Conservative 101; Mismatches 214; Indels 157; Gaps 25;

```

QY 19 VIQCKESHSTSRGELAINIDEPERSIDPROVLLSEISLVKHYEGIVENNLSGNI 78
Db 2 VLAACSGSSGSAKGEKTFYSIYETDPDNLNLTAKATANTISNVVDGL-ENRIGNF 60
QY 79 EPALAEYSLSDGLTYTEKL-KSAFW--SNQD--PLTADPIESKQVAVQEVSGIYA 132
Db 61 VPSMAEDSVSKDILITTYITRKDAKWYISGEYAAKADQFVGLTKYAAOKKSDALYP 120
QY 133 FALNDIKVNRIGQHL-SIDHGVHSPNESTLVLTLESPHSPLKLLAPPEFVHKSQ 191
Db 121 -VQESIKGLDAYVYGEIKDFQVIGALDEQYQYTLNKPEFSWNKSTMGVLAVY--NE 177
QY 192 RTLOSK-----SLPIASGAPYPAKNIKOKWIKLSNPHYVQSOVERKTTITHTI 241
Db 178 EPLNKGDPDFAKATDSSSLYNGPYLLKRSIVYKSSVEFAKNPNYMDKDHIDKYKLSFW 237
QY 242 --PDANTAACLFNQKLNMOGPPWGERI-----PQETL-----S 273
Db 238 DGQPTSKRAENFKDGLT-----AARLYPTSASFALSKSMKDNIVTQDSITLYLVGT 291
QY 274 NLQSKGHLHSFDVAGTSMITFNINKFPLNNKLEALASALDEALVSTI--FLGRAKTA 331
Db 292 NIDRQSYKYT---SKTSDQKASTKKALLNKDFQALAFQFDRTAVASQLNQICASKYL 348
QY 332 DHL-PTNHS-----YPERHQEQAQOAY 356
Db 349 RNLDVPTFYQADGKNFGDVKELVYTGDEMDKYNLADSDGLYNPKAKAFAPAKA--- 405
QY 357 AKLLEKALEBELOITAKDEHLNLEFPVSSASLSL--VOLLIREQKESLG- 405
Db 406 ---KSAQABEVYTP---IHLDPVDQATATKYVORVQSKSLIETLADNVITIDQ 456
QY 406 -----FALPIVKEFALLQADSSGNSFIATNGGMFADFPADPMAPL----- 445
Db 457 QLOKDEVNNTYFAENVAAGEDM-----DLSD-----NNGMCPDADASTYLDIDIRKSVG 505
QY 446 ---TIFAYPSG---VPPYALNHKDFLEIQNI-EQGDHOKRSELYSQASLYLEFFHII 497
Db 506 ESTKTYLGFDSGSDVNAKKVGLDYREKLYTEAGDETTDVAKRYKYAAQAWLTDSSALI 565
QY 498 EP 499
Db 566 IP 567

RESULT 4
US-08-600-993A-47
; Sequence 47, Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,993A
FILING DATE: 1-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 643 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-600-993A-47

Query Match          9.9%; Score 270.5; DB 2; Length 643;
Best Local Similarity 21.6%; Pred. No. 5.4e-17;
Matches 130; Conservative 101; Mismatches 214; Indels 157; Gaps 25;

19 VLQCKESHSSSTSRGELAINIRDEPRSLDPQVRLSEISLVKHIEGVLQVENNLSGN 78
2 VLAACSGSSAKGKETSIEYEDPNLWYLTAAATANTSVNDGL-ENDRYGN 60
79 EPLAEDYLSDDGLTTFKL-KSAFY--SNGD--PLTAEDEFSMKQVATQEVSGIYA 132
61 VPSNADMSVSKDGLTYTIRKDAKWYSEGEVAAVQAQDFVGLKYADKKSDALYP 120
133 FALNPLKVRKIOEGHL-SIDHFGVHSPNSTLVLTLESPTSHFLKLLALVPFPYHKSQ 191
121 -VQESIKGIDAVYKEIWDPSQVGRALDQVYOTLNKPSFPMNSKTTTGVLAHV--NE 177
192 RFLQSK-----SLPIASGAFYFKNIKOKWIKLSKNPHYNSQVETKTITIHFI 241
178 EFLNSKGDFFAKATDPSLSLYNGPYLTKSTIVTKSSVEFAKNPNWMDKNHIDKVLSTW 237
242 --PDANTAKLFGNCKLWMOGPPWGERI-----POETL-----S 273
238 DGDQTSKPAENKRGDSL-----AARLYPTSAPAELEKSMKDNIVYTQDSITYLVGT 291
274 NLSKGLHSPVACTSWLTPNINKPPLNMKLRALASALDKALVSTI--FLGRAKTA 331
292 NIDROSKYKT---SKTSDEOKASTKRALLNKDFRQALAFGPDRTAVASQLNGQTGASKIL 348
332 DILL--PTNIHS-----YEHOKOMARQAY 356
349 RNLFPVPTFVADGKNFGDWVKEKLYTYDEKKNVNLSDODGLYNPEKKAKAEFAKA--- 405
357 AKTLREALEELQITAKDELHLILPVSASSASLL--VOLIREOMKESIG----- 405
406 ----KSLQAEGVTFP---IHLDHPVDQATATTKVQVQSMKQSLAEATIGADNVLIDIQ 456
406 ----FAIPVKEFALLADLSSGNEFLAGCFADFPADPAFTL----- 445
457 QLOKDEVNNTTYFAENAGEDW-----DLSL-----NVGQDPEDFDPSTYDIITKPSVG 505
446 ----ITFAIPSG---VPPVAINHKDFLELLONI-EQEQDHQKRSELSVQASLYLETFHILI 497

```

```

Db 506 ESTKTYLFDSGENVAAKAVGLYDEKLYTAGDETTIDAKKIDKRAAQAMLTDSLTI 565
QY 498 EP 499
Db 566 IP 567

RESULT 5
US-09-385-028-9
Sequence 9, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwame A. Aidoo
APPLICANT: Ashish S. Paradkar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jenner Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P574520S2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 39305350
TELEFAX: (202) 39305350
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-385-028-9

Query Match          6.0%; Score 165.5; DB 4; Length 556;
Best Local Similarity 21.0%; Pred. No. 4.2e-07;
Matches 97; Conservative 74; Mismatches 212; Indels 79; Gaps 17;

34 GELAINIRDEPRSLDPQVRLSEISLVKHIEGVLQVENNLSG-----NIEPALAEYSIS 89
29 GILRIVRTDDPSLDPGNTYYATWNLRLIGRTVTFPTAPGAGORLVPDLAESIGES 88
90 S-DGLTYTTRKLSAR-WSNGDPLTAEDEFSMKQVATQEVSGIYAFALNPLKVRKIOEG 147
89 SEGGRWYTRLEGRYEDGTPVVSADI---KHAJARSNVGTIVLAGP-----TYRR 138
148 HLSIDHFG-----VHSPNSTLVLTLESPTSHFLKLLALVPFPYHKSQRT- 193
139 HLLGTGCGPWRPEADQSPVTLTETPDKTIVRLRLNRPFRAGMDLAIWSTTPVPRDRDYG 198

```


QY 260 GPPWGER-----IPQETLSMLQS--KGHLSPDVAGISWLTFFNINKFPLNNKL 306
 Db 425 KLPVYSRFLATNSKYMPEIYSEFINYLNDGFRSQHLS-----NKINVKNI-- 469
 QY 307 REALASALDKKALVSTYIFLG---RAKTAHLLPTNHSYF---EHOKQEMARQAYAKL 360
 Db 470 --IFSEMIKQQLIPSMIFHKIRLIMQVNNVEILTVLLEHSGKFLNKPXY----- 523
 QY 361 FEALAELOITAKDLEHLNLPVSSASSLLVOLIREQWESIGFAIPYQKE-----FA 416
 Db 524 -KELMEKVVQLIKDKNRDLNNKMSALENITLILPSPVKSILNTVKTITPEQGFYRI 582
 QY 417 LLOADLSSGNF---SLATGMPADPADMAFLTFAYP-----SGVPPY 457
 Db 583 LRSELSLDEKHALVKLYRAHMDVAIQKVLFSKPKHISYQNIPLITVLGGLSY 642
 QY 458 AINHKDFP-----EILONIE---QEDDHOKSELYSASLYLTFH-----ITPEY 501
 Db 643 ---RDFVIRICIDVLENIRGLEINDYGQNNHRISNVRYLTETIFEMIKSDVLDITLY 699
 QY 502 H 502
 Db 700 H 700

RESULT 8

US-09-177-431-2
 ; Sequence 2, Application US/09177431
 ; Patent No. 6071700
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobson, Allan S.
 ; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: PASTESQ for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/177,431
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/955,472
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fasse, J. Peter
 ; REGISTRATION NUMBER: 32,983
 ; REFERENCE/DOCKET NUMBER: 07917/050001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-9806
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1089 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-177-431-2

Query Match 3.9%; Score 107; DB 3; Length 1089;
 Best Local Similarity 18.6%; Pred. No. 0.46;
 Matches 112; Conservative 89; Mismatches 206; Indels 194; Gaps 30;

QY 52 VLLSEISLY-----KATIEGVQENNLSGNIPLALADYSLSDGLTYTFKL 99
 Db 144 LRVFTLYLVGFRTLDIESKDAIPNFLOKK--IGRMDPLL--FSILREIILNYFKILG 198
 QY 100 -----KSAPFWSNGDPLTAEDFIESMKQV-----AQE-----VSGIAPALNITKRV 141
 Db 199 FTTITATIKKFAFLRBD--DNSMDLLYDSKGLAQSLFKNFIDATFARFTELHKRV 257
 QY 142 RKIQEGHLS-----IDHFGVSHNESTLVLTLESPTSHPLKLLALVFPVH- 188
 Db 258 NKLOREHOKCOLRTGKLEDEVEEY-----DKLPIPIRRTKS-----ATLLEEFKLEI 307
 QY 189 ---KSQRTLOSKSLPIASATFPKNIKQKM-----IKLSKNPHY 225
 Db 308 PELEGASNDLKEFASPMITNQLPPN--QRLMENDYRKFYEILPDISKVEESQSKT 365
 QY 226 YNOSQVEYKTLTIHF-----IPD-----ANTAALFNOGKLKMQ 259
 Db 366 EKDSNNVSKNINLPFTDLEMAQCKDIIDLSNRVSSYLDNKAATRNRIKFFMETQ-DMS 424
 QY 260 GPPWGER-----IPQETLSMLQS--KGHLSPDVAGISWLTFFNINKFPLNNKL 306
 Db 425 KLPVYSRFLATNSKYMPEIYSEFINYLNDGFRSQHLS-----NKINVKNI-- 469
 QY 307 REALASALDKKALVSTYIFLG---RAKTAHLLPTNHSYF---EHOKQEMARQAYAKL 360
 Db 470 --IFSEMIKQQLIPSMIFHKIRLIMQVNNVEILTVLLEHSGKFLNKPXY----- 523
 QY 361 FEALAELOITAKDLEHLNLPVSSASSLLVOLIREQWESIGFAIPYQKE-----FA 416
 Db 524 -KELMEKVVQLIKDKNRDLNNKMSALENITLILPSPVKSILNTVKTITPEQGFYRI 582
 QY 417 LLOADLSSGNF---SLATGMPADPADMAFLTFAYP-----SGVPPY 457
 Db 583 LRSELSLDEKHALVKLYRAHMDVAIQKVLFSKPKHISYQNIPLITVLGGLSY 642
 QY 458 AINHKDFP-----EILONIE---QEDDHOKSELYSASLYLTFH-----ITPEY 501
 Db 643 ---RDFVIRICIDVLENIRGLEINDYGQNNHRISNVRYLTETIFEMIKSDVLDITLY 699
 QY 502 H 502
 Db 700 H 700

RESULT 9

PCT-US95-16930-2
 ; Sequence 2, Application PC/TUS9516930
 ; GENERAL INFORMATION:
 ; APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
 ; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE
 ; TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF
 ; TITLE OF INVENTION: NONSENSE-MEDIATED MRNA DECAY
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street Suite 3100
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/16930
 ; FILING DATE: 27-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/375,300

FILING DATE: 20-JAN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 04/20/046W01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)542-5070
 TELEFAX: (617)542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1089 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-16930-2

Query Match 3.9%; Score 107; DB 5; Length 1089;
 Best Local Similarity 18.6%; Pred. No. 0.46; Indels 194; Gaps 30;

Matches 112; Conservative 89; Mismatches 206; Indels 194; Gaps 30;

QY 52 VRLSEISLV-----KHIEGLVOENNLSCNIEPALAEDYSLSDELTYTEKL- 99
 Db 144 LRVETELVGVERTLDIESKDAIPNFKKK--TGRKDPPL--FSILRELIANKFKLG 198
 QY 100 -----KSAFNSGNDPLTAEDFIESKKQV-----ATQE-----VSGIYAPALNPKNV 141
 Db 199 FTTTIAIAFTAKKAPLFRPD-DNSMDLIDYSKIKGALQSLPKNFIDATFARATLHKV 257
 QY 142 RKIOEGHLS-----IDHGVHSPNESTLVVLTESPTSHFKLLLPFFPVH- 188
 Db 258 NKIQREHQKQIIRTKGLREDEYEEV-----DKLLPFRKFTS-----AIIIGFFPLEI 307
 QY 189 -----KQRILOSKSLPIASGAFYKNIKOKQV-----IKLSKNPHY 225
 Db 308 PELGASNDLKTASPMITVQILPPN--QRLWENEDTKRFEYELPIDISKIVESQSKPT 365
 QY 226 YNOSOVETKTIITHF-----IPD-----ANTAKILFNQKLNWQ 259
 Db 366 EKDSNVSKNINLFTDLMAKCDIIDDLSNKNYSYLDKNKATNRRLKFFMETQ-DWS 424
 QY 260 GPPVGER-----IPOETLSLQS--KGLHSFVASTSWLTFNINKEFLNNMKL 306
 Db 425 KLPVYSRFTATNSKYMPETVSEFTNYLDNGFRSOLHS-----NKIVAKNI-- 469
 QY 307 REALASALKEALVSTIFLG--RAKADHLPLTNHSTP--EHOQEMAQROAVAKKL 360
 Db 470 --IFFSEMIKPOLIPSEMIHFKIKITLIMYQVFNNEILLYLLEHSGFLINKREY---- 523
 QY 361 FKALBELOITAKDLEHNLIFPVSSASSLVOLLREQWKESLGFALPIYQKE----FA 416
 Db 524 -KEIMKMWOLLKDKKNDROKNNMKKALENITITLLIPPSYKSLNVVTKITTPEOQFPR 582
 QY 417 LLOADLSSGNF-----SLATGGEADPADPMAFLTIFAVP-----SGVPY 457
 Db 583 LINSSELSDLEFKHIVKLVRAHMDVAIOKVLFSKPHKISYQNTPLITKVLGSLY 642
 QY 458 AIHMKDFL-----EILQNI-----QEDHOKRESEIYQSLSLETFH-----IIEPLY 501
 Db 643 ---RDFVYRCIDOVLENIERGELINDYGNMHRISNVRLTEIFNEMIKSDVLDITLY 699
 QY 502 H 502
 Db 700 H 700

RESULT 10

us-08-328-256-10
 ; Sequence 10, Application us/08328256
 ; Patent No. 5643749
 ; GENERAL INFORMATION:

APPLICANT: REVEL, Michel
 APPLICANT: ABRAMOVICH, Carolina
 APPLICANT: RATOYITSKI, Edward
 TITLE OF INVENTION: SOLUBLE INTERFERON ALPHA-RECEPTOR, ITS
 PREPARATION AND USE
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEWMARK
 STREET: 419 Seventh Street, N.W., Suite 300
 City: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/328,256
 FILING DATE: 24-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IL 107378
 FILING DATE: 24-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: BROWDY, Roger L.
 REGISTRATION NUMBER: REVEL-13
 REFERENCE/DOCKET NUMBER: 25,618
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 557 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-328-256-10

Query Match 3.8%; Score 104.5; DB 1; Length 557;
 Best Local Similarity 19.3%; Pred. No. 0.26;
 Matches 114; Conservative 77; Mismatches 150; Indels 251; Gaps 33;

QY 18 VILGCKESHSSSTSGELAINIRDEPR-----SLDPRQV 52
 Db 74 IKISGCONITSTCNSSSLKLVYEIEIKLRIRAKENTSSWEVDSFPPFKAQIGPPEV 133
 QY 53 RLSE-LSLVKHIYEGLVQENNLSCNIEPALAEDYSLSDELTYTEFKLSAFNSGNDPLT 111
 Db 134 HLEAEDKAVIYH-----ISPGRKDSYVMAWDGLSFTYSL--LIMKNS----- 173
 QY 112 AEDTIESKQVAAQVSGIYAFALNPKNVKKTQEGHLSIDHGVHSPNESTLVVLTESP 171
 Db 174 -----SGVE-----ERIENTYSHR-----KIKYLSPE 195
 QY 172 TSHPLKLLA-----LPVFPVHKQSQRTLOSKSLASGATYFKNK----- 212
 Db 196 TTYCLAKAKALLSMKIGIYSPVHCITVENEPLP-----PENIVSVQONOVYLVK 248
 QY 213 -----OKWIK--LSKNP--HYNOSOVETKTIITHFIPD--ANTAKILFNQK 256
 Db 249 DYTANMTFOVQWMLHAFLEKRNPNHLYKWKQ-----IPDENKTKIQCYFPQ--- 295
 QY 257 NMOGPPWGERIPQETLSLQSKG-----HLHSDVASTSWLTFNINKEF-----PLN 302
 Db 296 -----NVFQKGIYLLRVQASDGNNTSFWSSEI--KEDTEIQAFLLPVE 337
 QY 303 NMLREALASALDKALVSTIFELGRATADHLPLTNHSTP--EHOQEMAQROAV 356
 Db 338 NLR---SLDSFR-----IYIGAPKQSN--TFV--IDYPLIYIITWENTSNAERKII 385

QY 357 AKKLEKLELELTAKDLEHLNLFPPVSSASSLLVLIQWKESLGLAIPYKGEFA 416
 Db 386 EKK-----TDVTPNLKPLT-VYCVARAHTM-----DEKLNKSVFSNVAEKEK--- 428
 QY 417 LLAADLSSGNFS---LATGGMFADFADPM-----AFLTIFAY---PSGVPPYAINH--- 461
 Db 429 -----TKPGNTSKTWLIVGICIALFALPVIYAAKVFILRCINIVFPSPSLKSSSIDEXFS 483
 QY 462 -----KDFL-----EILQNT-----EEDODHOKRSELVSQAS 488
 Db 484 EQLKNNLLSTSEQIEKCFIENISTLATVEETNQDDEDHKKYSSQTSQDS 535

RESULT 11
 US-08-471-454-2

Sequence 2, Application US/08471454

Patent No. 5731169

GENERAL INFORMATION:

APPLICANT: MOGENSEN, Knud E.

APPLICANT: UZE, Gilles

APPLICANT: LUTFALLA, Georges

APPLICANT: GRESSER, Ion

TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR

TITLE OF INVENTION: THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE

TITLE OF INVENTION: PREPARATION OF THE CORRESPONDING PROTEIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHAYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,454

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

NAME: BYRNE, THOMAS E.
 REGISTRATION NUMBER: 32,205
 REFERENCE/DOCKET NUMBER: 960-7
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 TELEX: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 557 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-466-974-2

Query Match 3.8%; Score 104.5; DB 2; Length 557;
 Best Local Similarity 19.3%; Pred. No. 0.26;
 Matches 114; Conservative 77; Mismatches 150; Indels 251; Gaps 33;

```

QY 18 VLQGGKSSHSSTSGELAINIRDEPR-----SLDPRGV 52
D 18 VLQGGKSSHSSTSGELAINIRDEPR-----SLDPRGV 52
D 74 IKLSGQNTSTKCNFSSLKLVYEIKLIRAKENTSSWEYDSFTPRKQIGPPEV 133
QY 53 RLISE-ISLVKHIYGLVQENNLSGNTEPALADYSLSSDGLTYFFKLKSAFMSNGDPLT 111
D 53 RLISE-ISLVKHIYGLVQENNLSGNTEPALADYSLSSDGLTYFFKLKSAFMSNGDPLT 111
D 134 HLEADKALVIH-----ISPQKDSVMMALDGLSFTYSL-LIKMNS---- 173
QY 112 AEDFISWQVATQVSGIYAFALNPKNVRIQEGHLSIDHFGVHSPNESTLVLTLESP 171
D 112 AEDFISWQVATQVSGIYAFALNPKNVRIQEGHLSIDHFGVHSPNESTLVLTLESP 171
D 174 -----SCGE-----ERLENISRH-----KIYLSPE 195
QY 172 TSHFLKLA-----LPVPPYKKSQRLQSKSLPIASGAFYRNK----- 212
D 172 TSHFLKLA-----LPVPPYKKSQRLQSKSLPIASGAFYRNK----- 212
D 196 TTYCLVKKALITSMKIGVSPVHCITVTVENLPP-----PENIEVSQONQYVLKW 248
QY 213 -----OKQWIK--LSKNP--HYNOSQVETITITIHPIP--ANTAALFNQGL 256
D 213 -----OKQWIK--LSKNP--HYNOSQVETITITIHPIP--ANTAALFNQGL 256
D 249 DYTANMTFOYQWHLAFKLRPNQNHLYKKQ-----IPDCENVKTKQCVFPQ--- 295
QY 257 NMGGPWWGERIPEQTLNLSQSG-----HLHSPDVAGTSMLEFNINKE-----PLN 302
D 257 NMGGPWWGERIPEQTLNLSQSG-----HLHSPDVAGTSMLEFNINKE-----PLN 302
D 296 -----NVFOKGIYLLRVQASDGNNTSFWSEEL-KFDEIQAFLPPVF 337
QY 303 NMLREALASALDEKALVSTIFLGRAKTDHLLPNIHSYP-----EHQKQMAQRAY 356
D 303 NMLREALASALDEKALVSTIFLGRAKTDHLLPNIHSYP-----EHQKQMAQRAY 356
D 338 NTR-----SLSDSFH-----ITIGAPKQSGN-TPV-IDYDPLIYEITFWENTSNAERKI 385
QY 357 AKKLFKALDELQITANDLEHMLIPVSSASLSLYQLIREOKKESIGRAIPYKERA 416
D 357 AKKLFKALDELQITANDLEHMLIPVSSASLSLYQLIREOKKESIGRAIPYKERA 416
D 386 EKK-----TDVTYVNLKPL-VYCVAKRAHTY-----DEKLINSSVFSADVCEK--- 428
QY 417 LIQADLSSGNF---LATGGWFADPADPM-----AFLITFAY---PSGVPPAIYNH--- 461
D 417 LIQADLSSGNF---LATGGWFADPADPM-----AFLITFAY---PSGVPPAIYNH--- 461
D 429 -----TRPGNTSKIMLIVGICIALEFALPEVIAAKVFLRCINVEFPRLKPSSSIDEXFES 483
QY 462 -----KDFL-----EILONT-----EORQHQKSELVSQAS 488
D 462 -----KDFL-----EILONT-----EORQHQKSELVSQAS 488
D 484 EGPLKNLLISTSEQIEKCFIENISTATVETWQDEBDEHKYSSQTSQDS 535

```

RESULT 13
 us-08-471-453-2
 : Sequence 2, Application US/08471453
 : Patent No. 5886153
 : GENERAL INFORMATION:
 : APPLICANT: MOGENSEN, Knud E.
 : APPLICANT: UZE, Gilles
 : APPLICANT: LUPPALLA, Georges
 : APPLICANT: GRASSER, Ion
 : TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR
 : THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE

TITLE OF INVENTION: PREPARATION OF THE CORRESPONDING PROTEIN
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHIVE P.C.
 STREET: 1100 NORTH GLEBE ROAD
 CITY: ARLINGTON
 STATE: VIRGINIA
 COUNTRY: U.S.A.
 ZIP: 22201-4714
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,453
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/900,642
 FILING DATE: 15-JUN-1992
 APPLICATION NUMBER: FR 89/13770
 FILING DATE: 20-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: BYRNE, THOMAS E.
 REGISTRATION NUMBER: 32,205
 REFERENCE/DOCKET NUMBER: 960-7
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 TELEX: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 557 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-471-453-2

Query Match 3.8%; Score 104.5; DB 2; Length 557;
 Best Local Similarity 19.3%; Pred. No. 0.26;
 Matches 114; Conservative 77; Mismatches 150; Indels 251; Gaps 33;

```

QY 18 VLQGGKSSHSSTSGELAINIRDEPR-----SLDPRGV 52
D 18 VLQGGKSSHSSTSGELAINIRDEPR-----SLDPRGV 52
D 74 IKLSGQNTSTKCNFSSLKLVYEIKLIRAKENTSSWEYDSFTPRKQIGPPEV 133
QY 53 RLISE-ISLVKHIYGLVQENNLSGNTEPALADYSLSSDGLTYFFKLKSAFMSNGDPLT 111
D 53 RLISE-ISLVKHIYGLVQENNLSGNTEPALADYSLSSDGLTYFFKLKSAFMSNGDPLT 111
D 134 HLEADKALVIH-----ISPQKDSVMMALDGLSFTYSL-LIKMNS---- 173
QY 112 AEDFISWQVATQVSGIYAFALNPKNVRIQEGHLSIDHFGVHSPNESTLVLTLESP 171
D 112 AEDFISWQVATQVSGIYAFALNPKNVRIQEGHLSIDHFGVHSPNESTLVLTLESP 171
D 174 -----SCGE-----ERLENISRH-----KIYLSPE 195
QY 172 TSHFLKLA-----LPVPPYKKSQRLQSKSLPIASGAFYRNK----- 212
D 172 TSHFLKLA-----LPVPPYKKSQRLQSKSLPIASGAFYRNK----- 212
D 196 TTYCLVKKALITSMKIGVSPVHCITVTVENLPP-----PENIEVSQONQYVLKW 248
QY 213 -----OKQWIK--LSKNP--HYNOSQVETITITIHPIP--ANTAALFNQGL 256
D 213 -----OKQWIK--LSKNP--HYNOSQVETITITIHPIP--ANTAALFNQGL 256
D 249 DYTANMTFOYQWHLAFKLRPNQNHLYKKQ-----IPDCENVKTKQCVFPQ--- 295
QY 257 NMGGPWWGERIPEQTLNLSQSG-----HLHSPDVAGTSMLEFNINKE-----PLN 302
D 257 NMGGPWWGERIPEQTLNLSQSG-----HLHSPDVAGTSMLEFNINKE-----PLN 302
D 296 -----NVFOKGIYLLRVQASDGNNTSFWSEEL-KFDEIQAFLPPVF 337
QY 303 NMLREALASALDEKALVSTIFLGRAKTDHLLPNIHSYP-----EHQKQMAQRAY 356
D 303 NMLREALASALDEKALVSTIFLGRAKTDHLLPNIHSYP-----EHQKQMAQRAY 356
D 338 NTR-----SLSDSFH-----ITIGAPKQSGN-TPV-IDYDPLIYEITFWENTSNAERKI 385

```

QY 53 RLLSE-ISLVKHIYEGIVQENNLSGNIEPALAEEDYSSSDGLTYTFKLKSAFMSNGDPLT 111

978 GAQHGQSHGICLLDFAL-----ISLIMDNELFAETFFENNYKSFTHPLCVLENDLAKRFIGA 1032


```

QY 177 KLALPFFPVHSHQSTQLOKSLP-----IASAEYPR-NIKOKW-IRLSKNP- 223
Db 1033 KSSAISTSL-TNLRNRESSSQIPPTSTSNQFSAIGFPEYDELKFPVMDLKVTSVEGC 1092
QY 224 -HYNNOUYERTKIITHIPD-----ANTAA-----KLFNOKSLMWOPPMGEXI 267
Db 1092 PLVDLAKIQGVLESILYHIEHADTDDJLANNKFTLDIKIINOEYVO---ENDKRL 1147
QY 268 POELTSLMOSKGLHSPVACTSMWLPENIKF-PLNNKRLREALASALDEKALVSTIFLG 326
Db 1148 DDLRNNNSOSKRNKSCDNDSNAKTFPHVDARPEUSNENNRGAATMLGJSSIA----- 1201
QY 327 RAKTADHLPTNHSYPRHQOEAQORAYAKKIRKALELEJOTTADEHMLJFPVSS 386
Db 1202 -----ALEKLOCTLOD-----LYVKIKS 1219
QY 387 SASLLWOLIREQWKESTLGAIPDPLVGEKFEALLOADLSSGNRLATGSMFADPADPAEPI 446
Db 1220 S-----YQQLATRLPGLATRNCRKVYHMLQOPPOJSMALLMNG-SSDTLDRK-VT 1268
QY 447 IFAPSGVPRPAINKKPELILONTIOEQODH--QKRSEIVSQASLETFPHALIEPYIDA 504
Db 1269 EF-----QALKMTDYDIL-NTWIKYKLDHFITSKILVSNQD-WLQVSQJLESLSDS 1318
QY 505 ----FOFANRKLKSLGVPSTGVVD 525
Db 1319 LVALFNYPLHAFESNNVIASSSSOLD 1343

```


[illegible]

```

1429 GACTGG.....GATTATACAT.....1446
430 aThGlyGlyTyrPheAlaSPheAlaSPrometaAlaPheLeu....445
1447 .AATGCGTGGGGTCAGACTTGGCGATCCATCACTACTGATA 1455
446 .....ThiLePheAlaIlyPro 451
1496 TCATCAACCATCTGTAGAGAAAGTACTAAACATATTAGGGTTGAC 1545
452 SerGly.....ValProTyrAlaIleAsnHisLysAsPheLe 465
1546 TCAGGGGAGATTAATGATGCTGTAAAAAGTAGCTATATGACTACGA 1595
465 uGluIleLeuGlnAsnIle...GluGlnGlnGlnAspHisGlnLysArgS 481
1596 AAAATTGCTTACTGAGGCTGTGTAGACTACAGATGTTGCTAAACGCT 1645
481 erGluLeuValSerGlnAlaSerLeuTyrLeuGluThrPheHisIleLe 497
1646 ATGATAATACGCTGCAGCCCACTGCTTGCACAGATGCTGTTGATT 1695
498 GluPro 499
1696 ATTCCA 1701

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:us-08-600-993A-46
seq_documentation_block:
; Sequence 46 Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; CELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,993A
; FILING DATE: 1-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2019 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: R6
; IMMEDIATE SOURCE:
; CLONE: SPUR98
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1932
; US-08-600-993A-46

```

```

alignment_scores:
  quality: 270.50      length: 602
  ratio: 0.902
  percent similarity: 49.834  percent identity: 21.595

```

alignment_block:

US-09-824-567-2 x US-08-600-993A-46 ..

Align seg 1/1 to: US-08-600-993A-46 from: 1 to: 2019

```

19 ValLeuGlnGlyGlySerGluSerSerHisSerThSerArgGlyG1 35
||||| :|||: |||: |||: |||:
4 GTACTGTCAGACAGCTCTGTGATCAGGTCAACGCTAAAGGTGAGAAAC 53
||||| :|||: |||: |||: |||:
35 uLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArgGlnv 52
||||| :|||: |||: |||: |||:
54 ATTCTATACATTATAGACAGACACCTGATTAACCTCAACTATTGACAA 103
||||| :|||: |||: |||: |||:
52 alArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGlnGlyLeu 68
||||| :|||: |||: |||: |||:
104 CTGCTAAGCTCGACAGCAAAATATTACAGTAGCTGTGATGATGTTG 153
||||| :|||: |||: |||: |||:
69 ValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAs 85
||||| :|||: |||: |||: |||:
154 CTA..GAAATGATCGCTACGGAGAACTTGTCCGCTATGCGTGAGGA 200
||||| :|||: |||: |||: |||:
85 pTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeu...lyss 101
||||| :|||: |||: |||: |||:
201 TTGCTGTATCCAGAGATGATTTGATTAACCTTATCTATCCGTAAAG 250
||||| :|||: |||: |||: |||:
101 erAlaPheTyr.....SerAsnGlyAsp.....ProLeuThAla 112
||||| :|||: |||: |||: |||:
251 ATCCAAATGATATCTTGAAGGTGAGAAATACGGCGGCACTAAAGCT 300
||||| :|||: |||: |||: |||:
113 GluAspPheIleGluSerTyrPlyGlnValAlaThrGlnGluValSerG1 129
||||| :|||: |||: |||: |||:
301 CAAAGACTTTGTACAGAGACTAAATATGCTGCTATAAATAATACAGATCG 350
||||| :|||: |||: |||: |||:
129 yIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGln 146
||||| :|||: |||: |||: |||:
351 TCTTACCT...GTCAACAATCAATCAAGGGTGGATGCTTATGTA 397
||||| :|||: |||: |||: |||:
146 lncGlyHisLeu...SerIleAspHisPheGlyValHisSerProAsnGlu 161
||||| :|||: |||: |||: |||:
398 AAGGGGAATCAAGATTTCTCAAGATGAGATTAAGCTTGGATGTA 447
||||| :|||: |||: |||: |||:
162 SerThrLeuValAlaThrLeuGlnSerProHisSerHisPheLeuLys 178
||||| :|||: |||: |||: |||:
448 CAGAGGTTCAGTACACTTGAACAACAGAAAGCTTCTGGAATTCCTA 497
||||| :|||: |||: |||: |||:
178 uLeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLeu 195
||||| :|||: |||: |||: |||:
498 GACAACCATGGGTGCTGCTGCCCGAGTT...AATGAGAGATTTTGA 541
||||| :|||: |||: |||: |||:
195 lncLys.....SerLeuProIle 201
||||| :|||: |||: |||: |||:

```

```

542 ATTCAAAAGAGATGATTTTGGCCAAAGCTACGATCAAGTAGTCTCTTG 591
202 AlaSerGlyAlaPheTyrProLysAsnIleIysGlnIysGlnTrrIlely 218
592 TATAGGGCTCTTATTTGTAATCATTCATTCAGCAATCTCTGTGA 641
218 sLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValIleThrLysr 225
642 ATTTGGAAATAATCCGAACTACTGGATAGACAAATGCGATTTGACA 691
235 hIleThrIleHisPheIle.....ProAspAlaAsnThrAlaAlaLys 249
692 AAGTTAAATTTCTCATTTCTGGATGTCAGATACAGCAAAACCTCGAGAA 741
250 LeuPheAsnGlnIysLysLeuAsnTrrPngIlyProProIrrIlyLys 266
742 AACCTTAAAGATGGTACCTTACA.....CGACCTG 773
266 gIle..... 267
774 TCTTATCCACAAAGTGCAGATTTGGCAGACTGACAAAGATATAGAG 823
268 .....ProGlnGluThrLeu.....Ser 273
824 ACAATATTGCTTATCTCAACAAGACTCTATTACGTATCTAGTGGTACA 873
274 AsnLeuGlnSerLysGlyHisLeuHisSerPheAspValIleAllyHis 290
874 AATATTGACCGCTACGCTTATAATACACA.....TCTAAGCCAG 914
290 TrPLeuThrPheAsnIleAsnLysPheProLeuAsnMetLysLeuA 307
915 CGATCAACAAGGATCGATCAAAAGCGCTCTTAAACAAGATTTCC 964
307 TGGIAlaIleuAlaSerAlaLeuAspLysGlnAlaLeuValSerThrIle 323
965 GTCAAGCTATGCTTGGTTGGTTCATCCAGCCCTTACCTCTCAGTTG 1014
324 .....PheLeuGlyArgAlaLysThrAlaAspHisLeu.....Pr 336
1015 AATGCACAACCTGGACAGATTAATCTTGCTATCTCTTTGTGCACC 1064
336 cThrAsnIleHisSer..... 341
1065 AACATTTGTCAGACAGATGTAAAACTTTGGCATATGTCCAAAGAGA 1114
341 ..... 341
1115 AATTGTCACCTATGGGATGATGAGAGATGTTAATTCGACATTTCT 1164
342 .....TyrProGlnHisGlnLysGlnIleMetAlaGlnAr 353
1165 CAGGATGGCTTTACAAATCCAGAAAAGCCAGCTGATTTGCTTAAAGC 1214
353 gGlnAlaTyrAlaLysLysLeuPheLysGlnAlaLeuGlnIleLeuGlnI 370
1215 T.....AAATCGCTTACAGACAGAGAGTG 1240
370 LeThrAlaLysAspLeuGlnHisLeuAsnLeuIlePheProAlaSer 386
1241 TGACATTTCCCA.....ATTATTTGGATTCGCGATGGACCAG 1278
387 SerAlaSerSerLeuLeu.....ValGlnLeuIleArgGlnIleTrrPly 401
1279 ACAGCAACTACAAAGTTCAAGCGGTCCATCATATGAACAATCTCTGGA 1328
401 sGluSerLeuGly..... 405
1329 AGCAACTTTAGAGCTGATTAATGTCATTAATGATTCACAACTACAAA 1378
406 .....PheAlaIleProIleValGlyLys 413
1379 AAGACAGATTAACAATATATTTCGCGAATAATGCTGCTGGCGAA 1428

```

```

414 GluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAl 430
1429 GACGCG.....GATTTATCAGAT..... 1446
430 aThrIlyGlyTrrPheAlaAspPheAlaAspPheAlaPheLeu.... 445
1447 .AATTCGGTGGGTGCGACACTTGGCGATCCATCAACCTACCTGTGATA 1495
446 .....ThrIlePheAlaTyrPro 451
1496 TCATCAAAACCATCTGTAGAGAAAGTACTTAAACAATATTAGTTAGTTGAC 1545
452 SerGly.....ValProProTyrAlaIleAsnHisLysAspPheLe 465
1546 TCAGGGGAAGATAATGTGCTGCTTAAAAAAGTAAAGTATATAGTACAGA 1595
465 uGluIleLeuGlnAsnIle...GluGlnGlnGlnAspHisGlnLysArg 481
1596 AAATTTGTTACTGAGCTGGTGTATGAGACTACAGATGTGCTTAAAGCT 1645
481 erGluLeuValSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIle 497
1646 ATGATTAATTCGCTGCAGCCCAAGCTTGGTGCAGATATGCTTTGATT 1695
498 GluPro 499
1696 ATTCCA 1701

```

```

seq_name: /cgn2_6/producta/2/ina/6B.COMB.seq:US-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
Quality: 176.00 Length: 528
Ratio: 0.649 Gaps: 21
Percent Similarity: 51.326 Percent Identity: 20.265

```

```

alignment_block:
US-09-824-567-2 x US-09-103-840A-2/rev ..

```

```

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

```

38 ILeAsnIleArgAspGluProArgSerLeuAspProArgGlnValArgLe 54
4099170GTCAACGGCGGGGAACGCCAACCCGCTGATCCAGACGGGACCAAGA 4099121
54 uLeuSerGluIleSerLeuValLysHisIleTyrGlnGlyLeuValGlnG 71

```

```

4099120CAGCAAGCGGGCCATCATCGATCGTTGTCGCCGGCTGATGTCCT 4099071
71 lubsansleuSerglyAsn1legluprolaLeuAlaIuAspTyrSer 87
   :: ::::: |||::: |||
4099070ATGACGCGCTGGC.....AAGCGTCGCTGGAGGTCCGCGACGA 4099030
88 leu...SersAspGlyLeuThrTyrThrPheLysLeuLysSeraLaph 103
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4099029ATGAGGCGCGATACGTCACACTCGGATCTCTCAACCCGGCTG 4098980
103 e...TTPserAsnGlyAspProLeuThrAlaGluAspPheLeglUsert 119
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098979GAGTTCACCGACGCGCTCGCGGTGAGCGCCACTGTTCCGACGCGT 4098930
119 rPlysGlnValAlaThrGlnGluValSerglyTLeuAlaPheAlaLeu 135
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098929GAGACTACGACCCCTGACACCAATGCCAATCGACACACTTTTTC 4098880
136 AspProTLeuLysAsnValArgLysIle.....GlnGluG1 147
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098879AGCCCGATGAGGGGTTCGATGACGTGGCGCGCCAGCGGATAGAG 4098830
147 yHisLeuSer1LeasPHisPheGlyValHisSerProAsnGluSergThrL 164
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098829CCGACCAACATGTC.....GGTTGGCGCTGTCACACACTGGAGT 4098786
164 euValValThrLeuGluSergProThrSeraHisPheLysLeuLeuAla 180
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098785TCAACGCTGAGCTTAAGCGCCGACCATGACTCATCGTTCGCGCTGGC 4098736
181 leuProValPhePheProValHisLysSergin...ArgThrLeuGlnSe 196
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098733CACAGCTGCTTTATCCGCTGCGCGACTCGCATTTCCGGACATGGCCG 4098686
196 rLysSerLeu...ProTLeaLaserGlyAlaPhe..... 206
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098685GTTTCGCGCAACCCGATCGGACGCGCCCTGACAAATCGCGCAGCGCC 4098636
207 .....TyrProLysAsn1LeLysGlnLysGlnTrrLeyLys 218
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098635CAGCGCGCGCTGGGAACACACAGTCAGA.....ATCGAC 4098598
219 leuSergLysAsnProHisTyrTyrAsnGlnSerginValGluThrTyrsh 235
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098597CTGTGCGCCACCCGACTNCCACGCAACCGCAAGCCCGCTTACAAAG 4098548
235 rLleThrLleHisPheLleProAspAlaAsnThrAlaAlaLysLeuPheA 252
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098547TTTGGCATTCAGTTCTACGCCAATCGACACCGCTATGCCGACTTGC 4098498
252 snGlnGlyLysLeuAsnTrrGlnGly.....ProProTrrGlyGluArg 266
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098497TGTCTCGGCAATCTCGATGTCCTGGACAGATTCGCGACCGCTGAGC 4098448
267 lIleProGlnGluThrLeuSeraSnLeuGlnSergLysLysLeuHisSe 283
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098447GTCACACAGCGCACTGGCGCACACACCTTCCACGCGCGCCGACGAT 4098398
283 rPheAspValAlaGlyThrSerrPleuThrPheAsn1LeasLysPheP 300
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098397CAATTCAGACCCCTGACACT.....CCGTTCCGCGCGCGCATTTGC 4098357
300 rOleuAsnAsnMetLysLeuArgL...AlaLeuAlaSerAlaLeuAsp 315
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098356GTGGCGAGGCGCGCTGCGCGGTGGCGCTGTCGCCGCCCATCAAC 4098307
316 LysGlnAlaLeuValSerrThrLlePheLeuGlyArgAlaLysThrAla.. 331
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098306CGCGCAATCTGCCACAGATCTCGCGCGAACCAGCGGATCGGCGCG 4098257
332 .....AspHisLeuLeuAsnProThrA 338
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098256GATTTACCGCCGCTGTCGCGGCTTCGATTCACACCTGCGCGGCA 4098207

```

```

338 sn...LleHisSerrTyr...ProGlnHisGlnLysGlnGlnGlnGln 352
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098206ACGAGGTGCTGGACTACGACCCGACGCGGCTCGCGCTTGGCGCAG 4098157
353 ArgGlnAlaTyrAlaLysLysLeuPheLysGlnGlnGlnGlnGlnGln 369
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098156CGCATCGATCTCCACCGTGGAGCGCGCGGAGCGATC..... 4098118
369 nTleThrAlaLysAspLeuGlnHisLeuAsnLeuLlePheProValSers 386
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098117.GCCTATTAACGCTGCGCATCGGAGATGGGTGGAGCGCGGTGGCA 4098069
386 eSeraLaserSerrLeuLeu...ValGlnLeuLleArgGlnGlnTrrLys 401
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098068ACACATCAAGAACGTTGGGACATCATGCGGTGGCGCGCGCGACCC 4098019
402 GluSerrLeuGlyPheAlaTleProTLeValGlyLysGlnPheAlaLeu 418
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4098018ACCTTGGCGCGATTTGCGACCCAGATCAACACCGCGCATGAC.... 4097974
418 uGlnAlaAspLeuSerrSergLysAsnPheSerrLeuAlaThrGlyLysTrrP 435
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097973.....AGTGCATTTCCGCGCTGGATGCG 4097952
435 heLAspPheAlaAspProMetAlaPheLeuThr...LlePheAlaTyr 450
   :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097951GCGGCGATCAACCGCTGATGATGATGATGATGATGATGATGATGAT 4097902
451 ProSergLysValProProTrrAlaLleAsnHisLysAspPheLeuGlnL 467
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097901GCGCGCGGATCAACAGAGCTGCGCTTACATCAACCGCAATTTGACCGCGC 4097852
467 eLeuGlnAsn1LleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 484
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097851GCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4097802
484 aLserGlnLaserLeuLysLeuGlnThrPheHis.....LleLleGln 498
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097801TTCACACACCGCGCGCAATCTG.....TTTCACGACATGCCAGTTGTG 4097758
499 ProLleTyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSeraS 515
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097757CGCTGGGATTAACAGTGTGTCGATGTGTGTGTGTGTGTGTGTGTGT 4097708
515 nLeuGlyValSerrProThrGlyValValAspPhe 526
   ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
4097707GTTCACCGCTCACTGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4097674
seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

alignment_scores:
 Quality: 176.00 Length: 528
 Ratio: 0.649 Gaps: 21
 Percent Similarity: 51.326 Percent Identity: 20.265

alignment_block:
 US-09-824-567-2 x US-09-103-840A-1/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529

```

38 11easnlleargaspGluProArgSerLeuAspProArgGlnValArgLe 54
      ::::: |||::: |||::: |||:::
4106737GTCAACGGCGGGACCGCCCAACCGGCTGATCCGACCGCCACCA 4106924
      ::::: |||::: |||::: |||:::
54 uLeuSerGluLeuSerLeuValLysHisLleTyrGluGlyLeuValGln 71
      ::::: |||::: |||::: |||:::
4106923CACGACGAGTGGCGGATGATGATGATGATGATGATGATGATG 4106874
      ::::: |||::: |||::: |||:::
71 1uAsnAsnLeuSerGlyAsnLleGluProAlaLeuAlaGluAspTyrSer 87
      ::::: |||::: |||::: |||:::
4106873ATGACGCGCGTTGGC.....AGCGGTGCTGGAGGTGCGGAGTCA 4106833
      ::::: |||::: |||::: |||:::
88 Leu...SerSerAspGlyLeuThrTyrThrPheLysLeuLysSerAla 103
      ::::: |||::: |||::: |||:::
4106832ATCGAGCGCCGATACGTACATACGATGATGATGATGATGATG 4106783
      ::::: |||::: |||::: |||:::
103 e...TyrSerAsnGlyAspProLeuThrAlaGluAspPheLleGluSer 119
      ::::: |||::: |||::: |||:::
4106782GAATTCACCGAGCGTCCCGGAGGAGGAGGAGGAGGAGGAGG 4106733
      ::::: |||::: |||::: |||:::
119 rPlysGlnValAlaThrGlnGluValSerGlyLleTyrAlaPheAlaLeu 135
      ::::: |||::: |||::: |||:::
4106732GGAAGCTACGAGAGCCCTGAGACCAATGCGCACTGCAGAGCATTTTC 4106683
      ::::: |||::: |||::: |||:::
136 AsnProLleLysAsnValArgLysLle.....GlnGlnGlu 147
      ::::: |||::: |||::: |||:::
4106682AGCCCATCGAAGGTTGATGATGATGATGATGATGATGATGATG 4106633
      ::::: |||::: |||::: |||:::
147 yHisLeuSerLleAspHisPheGlyValHisSerProAsnGlnSerThr 164
      ::::: |||::: |||::: |||:::
4106632CCGACCATGTC.....GGGTGCGCGGTGTCACGACGTCGAGT 4106589
      ::::: |||::: |||::: |||:::
164 euValValThrLeuGluSerProThrSerHisPheLeuLysLeuAla 180
      ::::: |||::: |||::: |||:::
4106588TTCACCGTGTAAAGCGCCGACCATGCACTGACGTTGGGCTTGGC 4106539
      ::::: |||::: |||::: |||:::
181 LeuProValPhePheProValHisLysSerGln...ArgThrLeuGln 196
      ::::: |||::: |||::: |||:::
4106538CAGAGCTGTTTATCCGCTGCGGACTCGCATTTTGGGACATGGCCG 4106489
      ::::: |||::: |||::: |||:::
196 rLysSerLeu...ProLleAsnSerGlyAlaPhe..... 206
      ::::: |||::: |||::: |||:::
4106488GTTGCGCCGACCATGCGACGCGCCGATCAACAATCCGCCACGCC 4106439
      ::::: |||::: |||::: |||:::
207 .....TyrProLysAsnLleLysGlnLysGlnTyrLleLys 218
      ::::: |||::: |||::: |||:::
4106438CAGCGCGCGCCCTGGGAACACACAGTCAGA.....ATGAC 4106401
      ::::: |||::: |||::: |||:::
219 LeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrLys 235
      ::::: |||::: |||::: |||:::
4106400CTGTGCCCAACCCGACTACACGCGCAACCGCAACCCGCTAACAAAG 4106351
      ::::: |||::: |||::: |||:::
235 rLleThrLleHisPheLleProAspAlaAsnThrAlaAlaLysLeuPhe 252
      ::::: |||::: |||::: |||:::
4106350TTTGGATGAGTTTACGCCCATGTGACACCGCCATGCGGACTGTC 4106301
      ::::: |||::: |||::: |||:::
252 singleLysLeuAsnTyrGlnGly.....ProProTyrGlyGluArg 266
      ::::: |||::: |||::: |||:::
4106300TGTCCGCAATCTGATGTCGAGACAGATTCGCGGACGCGGTGAGC 4106251
      ::::: |||::: |||::: |||:::
267 lLeProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHis 283
      ::::: |||::: |||::: |||:::

```

```

4106250GTCTACACGCGGACCTGGCGCACCACTACAGCGGCGCCGACGAT 4106201
283 rPheAspValAlaGlyThrSerThrPheLeuThrPheAsnLleAsnLysPhe 300
      ::::: |||::: |||::: |||:::
4106200CAATCAAGACCTTCGACACT.....CCGTTGGCGGTGCGCATTTGC 4106160
      ::::: |||::: |||::: |||:::
300 rLeuAsnAsnMetLysLeuArgLys...AlaLeuAlaSerAlaLeuAsp 315
      ::::: |||::: |||::: |||:::
4106159GTGCGAGAGGCGGCGTGGCGCGGTGGCGGTGGCGGTGGCGGTG 4106110
      ::::: |||::: |||::: |||:::
316 LysGlnAlaLeuValSerThrLlePheLeuGlyArgAlaValSerAla 331
      ::::: |||::: |||::: |||:::
4106109CGCGCAAAATCTCGACAGATCTTCGCGGAGACCCGACATCGGCG 4106060
      ::::: |||::: |||::: |||:::
332 .....AspHisLeuLeuProThrA 338
      ::::: |||::: |||::: |||:::
4106093GATTTACACCCCGCGTGTGGCGGCGGTTCGATCCCACTCCGCGG 4106010
      ::::: |||::: |||::: |||:::
338 sn...LleHisSerTyr...ProGlnHisGlnLysGlnMetAlaGln 352
      ::::: |||::: |||::: |||:::
4106009ACGAGGTGCTGACTACGACCGCGACGCGGCTCGGCGGCTGAGG 4105960
      ::::: |||::: |||::: |||:::
353 ArgGlnAlaTyrAlaLysLysLeuPheLysGlnAlaLeuGlnGluLeu 369
      ::::: |||::: |||::: |||:::
4105959CGCATGCGATCTACCGGTGAGCGCGCGGTACCGGATC..... 4105921
      ::::: |||::: |||::: |||:::
369 nLleThrAlaLysAspLeuLysLeuAsnLeuLlePheProValSer 386
      ::::: |||::: |||::: |||:::
4105920.GCCATACACCTGACCTGCTGCTGGATGGATGGATGGATGGATG 4105872
      ::::: |||::: |||::: |||:::
386 eSerAlaSerSerLeuLeu...ValGlnLeuLleArgGlnGlnTyrPlys 401
      ::::: |||::: |||::: |||:::
4105871ACAGCATCAAGACGTTGGGATGATGGGATGGGATGGGATGGGATG 4105822
      ::::: |||::: |||::: |||:::
402 GluSerLeuGlyPheAlaLleProLleValGlyLysGluPheAlaLeu 418
      ::::: |||::: |||::: |||:::
4105821ACCTTCGCGGATTTCCGACCCGATCAACACCGCCCATGAC..... 4105777
      ::::: |||::: |||::: |||:::
418 uGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyLysTyr 435
      ::::: |||::: |||::: |||:::
4105776.....AGTGCATTTGCGCGGTGATGAGC 4105755
      ::::: |||::: |||::: |||:::
435 heAlaAspPheAlaAspProMetAlaPheLeuThr...lLePheAlaTyr 450
      ::::: |||::: |||::: |||:::
4105754GGGCGACTCCCGCTGATGATGATGATGATGATGATGATGATG 4105705
      ::::: |||::: |||::: |||:::
451 ProSerGlyValProProTyrAlaLleAsnHisLysAspPheLeuLle 467
      ::::: |||::: |||::: |||:::
4105704GGCGCGGATCCAGACGCTGCTACATCAACCCGGAATTCGACGCG 4105655
      ::::: |||::: |||::: |||:::
467 eLeuGlnAsnLleGluGlnGlnAlaAspHisGlnLysArgSerGluLeu 484
      ::::: |||::: |||::: |||:::
4105654CTGCGCGCGCGCGAAGCGCGCGCGCGGTACACCAATCCACGAG 4105605
      ::::: |||::: |||::: |||:::
484 aLleSerGlnAlaSerLeuTyrLeuGluThrPheHis.....lLeLleGlu 498
      ::::: |||::: |||::: |||:::
4105604TCAACAGCGCGGACGCAATCTCG.....TTTCACACATGTCATGTG 4105561
      ::::: |||::: |||::: |||:::
499 ProLleTyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSer 515
      ::::: |||::: |||::: |||:::
4105560CCGCTCGGATTTACATGATGTCGTGGATGTCGTGCTGCAGATGAC 4105511
      ::::: |||::: |||::: |||:::
515 nLeuGlyLysLeuSerProThrGlyValAlaAspPhe 526
      ::::: |||::: |||::: |||:::
4105510GTCACCGCTACCTGGAATGTCTGCGGACTGAC 4105477
      ::::: |||::: |||::: |||:::

```

seq_name: /cgn2_6/prodata/2/lna/6B.COMB.seq:US-09-385-028-20

seq_documentation_block:
 ; Sequence 20, Application US/09385028
 ; Patent No. 6232106
 ; GENERAL INFORMATION:
 ; APPLICANT: Susan E. Jensen

[illegible][illegible]


```

1 CLASSIFICATION:
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: US 08/790,462
4 FILING DATE: 29-JAN-1997
5 ATTORNEY/AGENT INFORMATION:
6 NAME: D. Douglas Price
7 REGISTRATION NUMBER: 24,514
8 REFERENCE/DOCKET NUMBER: 1418/0574520S2
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (202) 638-6666
11 TELEFAX: (202) 39305350
12 TELEX: RCA 248593 IDEA UR
13 INFORMATION FOR SEQ ID NO: 1:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 15079 base pairs
16 TYPE: nucleic acid
17 STRANDEDNESS: single
18 TOPOLOGY: linear
19 MOLECULE TYPE: DNA (genomic)
20 HYPOTHEtical: NO
21 ORIGINAL SOURCE:
22 ORGANISM: Streptomyces clavuligerus
23
24 US-09-385-028-1
25
26 Alignment scores:
27 quality: 165.50 Length: 462
28 Ratio: 0.710 Gaps: 17
29 Percent Similarity: 50.433 Percent Identity: 20.996
30
31 Alignment block:
32 US-09-824-567-2 x US-09-385-028-1 ..
33
34 Align seg 1/1 to: US-09-385-028-1 from: 1 to: 15079
35
36 34 GlyIuLdeAlaAlaAsnIleATGAspGluProAlySerLeuAspProx 50
37 ||| ||| ::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
38 9325 GGGAGCGCTGGCGTGGTGGCCAGAGACAGACTTCAGCTGGTCACCCCGG 9374
39
40 50 gGlnValArgLeuLeuSerGluIleSerLeuValIyHisIleTYGluG 67
41 ::::: ::::: ::::: ::::: ||| :::::
42 9375 CAACAGCACTACCCCTCACTGAGACTTCCTCGGCTCATCGGCGGA 9424
43
44 67 lYLeuValGlnGlnAsnLeuSerGly.....AsnIleGlu 79
45 ||||| ::::: ||||| :::::
46 9425 CGCTGTACCTTCGAGACACCGCGCGGCAAGCGGCGCTCTG 9474
47
48 80 ProAlaLeuAlaGluAspTYrIleSerLeuSer...AspGlyLeuThy 95
49 ||| ||||| ||||| ::::: ||||| ||||| ||||| |||||
50 9475 CCGGACCTCGCGGAGTCTGGGAGTCTCGAGAGCGCGGGGTCTG 9524
51
52 95 rThrPheIyAsnIleYsSerAlaPhe...ThpSerAsnGlyAspProLeuT 111
53 ::::: ::::: ::::: ::::: ::::: ||||| :::::
54 9525 GACCTACCGGCTCGGAGAGGCTTCGCTCAAGAGAGCGACGCGCG 9574
55
56 111 hralaGluAspPheIleGluSerTTPrySGlnValAlaThrGlnGluVal 127
57 ::::: ||||| ||||| :::::
58 9575 TCTGGCGCGACATC.....AAGCAGCCATCCCGCGAGAC 9612
59
60 128 SerGlyIleYrAlaPheAlaLeuAsnProIleYsAsnValArgIySII 144
61 ||| ::::: ::::: |||||
62 9613 TACGAGCACCATGCTCTGGGCGCGGTCCG.....AC 9644
63
64 144 eGlnGluIyGlnIleSerIleAspHisPheGly..... 155
65 ||||| |||||
66 9645 CTACTTCGCGCACTTCGGGACCGAGTACCGACCGCCCTGGCGGAGC 9694
67
68 156 .....ValHiserProAsnGluSerThleu 164
69 ||||| ||||| ||||| |||||
70 9695 CGAGCGCGGACGAGCGGCTGACGGTGGAGACCCGAGAGCGAGCGTg 9744
71
72 165 ValValThrLeuGluSerProThrSerHisPheIleuIyLeuAlaIe 181
73 ||| ||||| ||||| |||||

```

```

9745 GTCCTCCGGCTGCGGGAGCCGTTCCGGGGATGATCTCTGCGACCAT 9794
181 uProValPhePheProValHisLysSerGlnArgThr...LeuGlnSerL 197
:::|||||:::|||||:::
9795 GCGCTGCACACACCCCGCTGCCCGCGACCGGACACCGCGCGCATACC 9844
197 ySSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIleLysGln 213
:::|||||:::|||||:::
9845 GGCCTCCGGCGCGTGGCGACCGCCGCTGACCGGATGCTGTACACCCCG 9894
214 LysGlnTrrPileLysLeuSerLysAsnProHisTyrTrsGlnSerG1 230
:::|||||:::|||||:::
9895 GCGCAGCTGCGCTGCTGAGCCCAATCCGACCTGGAGCCGACACCGCA 9944
230 n.....ValGluThrLysThrIleThrIleHisPheIleProA 243
:::|||||:::|||||:::
9945 CCCGGTCCGGCTCCACCGCGCTCCCGGATGAGCTGACACCTGCGCAGG 9994
243 sPAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrrP1n 259
||:::|||||:::|||||:::
9995 ACCCGCACGAGTGGACCCGATGCTGCTGGCGGCGCAGCCCATGTGGAC 10044
260 GlyProProTrrPgly.....GluArgIleProGlnG1 270
:::|||||:::|||||:::
10045 CTCGCGGCGCTTGGTGTGCGAGCCGCGCGCCAGAGAGCGCATC..... 10086
270 uThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValA 287
||:::|||||:::|||||:::
10087 ....CTGCGGACCGGAGCTGCGCGCGCACCGCGCACAC...CCGCTGA 10129
287 laGlyThrSerTrrPleuThrPheAsnIleAsnLysPhe...ProLeuAsn 302
:::|||||:::|||||:::
10130 CCGGCTTACACTGATGATCTACTGCTGTCGAGCGCGGATCCCGCTTTCAC 10179
303 AsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLe 319
||:::|||||:::|||||:::
10180 AATGTGACTGCGCGCGCGCTGCTGCGACGTCGCCACCGCAAAAGCGGCAT 10229
319 uValSerThrIlePheLeuGlyArgAlaLysTrrAlaAsp.....H 333
:::|||||:::|||||:::
10230 GCAGGAGGCG.....TACGGCGCGCGGCGCGCGCGCACATCGCGACCA 10273
333 lSLeuLeuProThrAsnIleHisSerTyrProGlnHisGlnLysGlnG1 349
||:::|||||:::|||||:::
10274 CCGTGGCGCGCGCGCGCGCTGCGACGCTGACAGCATCTGACCGCTACCCG 10323
350 MetAlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuG1 366
:::|||||:::|||||:::
10324 GTCGCG.....CCGAGAGGACCGCG 10343
366 uGluLeuGlnIleThrAlaLysAspLeuGlnHisLeuAsnIlePheP 383
:::|||||:::|||||:::
10344 CGACCTGGAGGCGCGCGCGCGCGCGGAGCTGGCGCGGATGCCGACG 10393
383 roValSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGln 399
:::|||||:::|||||:::
10394 GCTTCCGACACAGATCGCGCGCGCGACAGACCGGCTCAAGAGATACCGG 10443
400 TrrPlyGlnSerLysLeuGlyPheAlaIleProIleValGlyLysGluPheAl 416
||:::|||||:::|||||:::
10444 GCGCGGAGCGCTGCGCGCGCGCTGCGCGGCTGCGCATCGAGGCGGA 10493
416 aLeuLeuGlnAlaAspLeuSerSerGlyAsnPe..... 427
:::|||||:::|||||:::
10494 GTGTCTG.....GACTTCGCGTGGCGGCGACTACTTCGACCGCTACGCG 10537
428 .....SerLeuAlaThrGlyGly 433
:::|||||:::|||||:::
10538 GCTGCCCGGAGATCTGGCGGACGACGAGGATCGGATCATATGTTGCGC 10587
434 TrrPheAlaAspPheAlaAspProMetAlaPheLeu 445
||:::|||||:::|||||:::
10588 TGGGGCGCGGACTTCCCGACGAGATACGCGCTTCTC 10623

```

```

seq_name: /cgn2_6/prodata/2/1na/PC/US_96-05320A-113
seq_documentation_block:
; Sequence 113, Application PC/US9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 22, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2232 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US96-05320A-113

alignment_scores:
Quality: 138.50 Length: 654
Ratio: 0.478 Gaps: 32
Percent Similarity: 44.343 Percent Identity: 18.807

alignment_block:
US-09-824-567-2 x PCT-US96-05320A-113 ..
Align seg 1/1 to: PCT-US96-05320A-113 from: 1 to: 2232
6 ValGlyIleGlySerThrIleLeuLeuSerLeuSerValIleGlnG1 22
:::|||||:::|||||:::
1 ATGATATTTTAAATTAAATAACGATATCTTTTATTTAGTTACCTTGAGG 50
22 yCysLys.....GluSerSerHisSerSerThrSerArgG 34

```

```

1  ::::: |||:|||||
51 TATCACTAATGTAACCGCAAAACAAATCTCTTAATTCTTAATGAG 100
34 IyGluLeu.....AlaIleasnIleArgAspGluProArgSerIeu 47
|||:||||| ::|:||||| ::|:|||||
101 AGGAGTTTATACCTATTATTGTGACACCTAATGATGATAGTAATAACTA 150
48 AspProArgIleValArgIleuLeuSerGluIleSerIeuVallys.... 62
|||:||||| ::|:||||| ::|:|||||
151 CTGGTCTGCTGATTAATAACAGAAATAATCGAACACACACAGCGGA 200
63 .....HisIleArgGluIleValGlnGlnAsnA 73
201 TAATCGACAAATTAATTAATTAATCTTGCCAGGTGTG.....A 238
73 snIleuSerGlyAsnIleGluPro..... 80
|||:||||| ::|:||||| ::|:|||||
239 ATATGGCTGGGGGTTCCGCCCTGGTGGCGCAACATTAATTAATGGA 288
81 ...AlaIleuAlaGluAspTyrSerIleuSerSerAspGlyLeuThrTh 96
::|:||||| ::|:||||| ::|:|||||
289 ATGGGTGATGCTGAAGATGTAGAGTGAAGTGAAGTGGCTTACAAAAAG 338
96 rPheLysLeu..... 99
|||:||||| ::|:||||| ::|:|||||
339 TTTTCAAAATACCAACAGGCTTCTATTATTGACCTGCAATTTTAA 388
100 .....LysSerAlaPheIlePheIlePheIle 107
|||:||||| ::|:||||| ::|:|||||
389 GAGAGTACACGTAGCAAGAAATTAATCTCTCAATATGTAATGTGT 438
108 .....AspProIleuThrAlaGluAspPheIle 116
|||:||||| ::|:||||| ::|:|||||
439 GGTTTTGGTGAACAGTAAATTAATTAACCTAGAGTGTAGGATTTCT 488
116 eolIleuSerTrpLysGlnValAlaThrGlnGluValSerGlyLeu..... 130
::|:||||| ::|:||||| ::|:|||||
489 GCAAGAA.....AACCAAGAAATAGGTGTTTATTAACT 523
131 .....TyrAlaIlePheAlaLeu... 135
|||:||||| ::|:||||| ::|:|||||
524 ATGGAATATATAGCAATATATACCAAAAAACCTTACTGACTAGCTAGT 573
136 .....AspProIleLysAsnValAlaArgLysIleGlnGluLysIleuSe 150
|||:||||| ::|:||||| ::|:|||||
574 TTACAATATGACAAAAATATGATTTATTACTATTGTGT.....TC 617
150 rIleAspHisPheGlyValHisSer...ProAsnGluSerThrIleuVal 166
::|:||||| ::|:||||| ::|:|||||
618 TGTAGAAATGCTGGCGATTATATAAGAACCGATATAGCAAAATCTTT 667
166 alIleuLysGluSerProThrSerHisPheLeuLys..... 177
|||:||||| ::|:||||| ::|:|||||
668 TTCAAAAAATATATCAAAAAACCGGATTATTAATTAATTAATGCAAT 717
178 .....LeuLeuAlaIleuProValAlaPhePheProValHisLysSe 190
|||:||||| ::|:||||| ::|:|||||
718 TCACCCGAGCATTTACTACATTTATGCGAGTGTATTATGTAATTAAG 767
190 rGlnArgThrLeuGlnSerLysSer.....LeuProIleAlaSerGlyA 205
::|:||||| ::|:||||| ::|:|||||
768 CTGGAGCGCTTTCGACAGCAAAAGATATCTCTAAACCAAGTTTGA 817
205 IapheIlePheLysAsnIleLysGlnLysGlnIlePheLysLeuSerLys 221
::|:||||| ::|:||||| ::|:|||||
818 GGTATATATATGCGTTATGTTACTGATATAGCGGTG.....AAA 855
222 AsnProHisIleTyrAsnGlnSerGlnValGlnThrLysThrIleThrI 238
::|:||||| ::|:||||| ::|:|||||
856 CGTAACCTGGTTTATGAGATCAAAAAAATATCAATTAATTAATAATA 905
238 eHisPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyL 255
::|:||||| ::|:||||| ::|:|||||

```

```

906 CAACTATCTACCTGAAAAATAACCATGATTAATTAATTAATCTACTAG... 951
255 yLysLeuAsnTrpGlnGlyProProTrpGlyGluArgIleProGlnGluThr 271
::|:||||| ::|:||||| ::|:|||||
952 ..TTTCAGCTACAGTAAACCAACCAAAATGATATGCTGCCAAAGAGCA 999
272 LeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValAlaG 288
|||:||||| ::|:||||| ::|:|||||
1000 TCTTCAGATTTAGTGGT.....ACCTAGGTAA 1028
288 yThrSerTrpLeuThrPheAsnIleAsnLysPheProLeuAsnMetL 305
::|:||||| ::|:||||| ::|:|||||
1029 TCAAACTGGAATCACTATTCAGATCTTACTTTGATATTAATAAC.... 1074
305 yLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSer 321
|||:||||| ::|:||||| ::|:|||||
1075 .....ACAAGTACTTTAATATCAAAACTACTGTTCAAT 1107
322 ThrIlePheLeuGly..... 326
|||:||||| ::|:||||| ::|:|||||
1108 GAACCTATTATTGTTGGTTTCAATGGTTAAAAACACAAAGAAATACTTAA 1157
327 .....ArgAlaLysThrAlaAspHisLeuLeuProThrAsnI 339
|||:||||| ::|:||||| ::|:|||||
1158 GTATGATTAAGTAAAGTGAAGAAAGCGAT..... 1188
339 IeHisSerTrpProGlnHisGlnLysGlnGluMetAlaGlnArgGlnAla 355
::|:||||| ::|:||||| ::|:|||||
1189 ..TATATATATGCGATATTTCACACCCATTAATGCAATCTGACAGTCGA 1236
356 TyrAlaLysLysLeuPheLysGlnAlaLeuGluGluLeuGlnIlePheAl 372
|||:||||| ::|:||||| ::|:|||||
1237 TATACACAGCGCTTTTATTACAA..... 1260
372 alYsAspLeuGlnHisLeuAsnLeuIlePheProValSerSerSerAla 389
|||:||||| ::|:||||| ::|:|||||
1261 .GATCAATTAATAATGAAAGATTAATTTTACGACAGCA..... 1299
389 eSerLeuLeuValGlnLeuIleArgGluGlnTrpLysGluSerLeuGly 405
::|:||||| ::|:||||| ::|:|||||
1300 .....GTGCGTTATGACCATATCAATTAATTAATGAGG 1329
406 PheAlaIleProIleValGlyLysGlnPheAlaLeuLeuGlnAlaAspLe 422
|||:||||| ::|:||||| ::|:|||||
1330 .....CAGAAAAATTTAGCACTTAATAATTAATGATAT 1361
422 uSerSerGly.....AsnPheSerLeuAlaThrGly 433
|||:||||| ::|:||||| ::|:|||||
1362 CTCTGAGGACATGATTAATGACCAAAAAATTAAT..... 1399
433 IyTrpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAla 449
|||:||||| ::|:||||| ::|:|||||
1400 GTTGG.....TCG 1407
449 TyrProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuG 466
|||:||||| ::|:||||| ::|:|||||
1408 TATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1448
466 uIleLeuGlnAsn.....IleGlnGlnG 474
|||:||||| ::|:||||| ::|:|||||
1449 TTTATTTACGAATTTTACTAAACTGGGAGACACCTGTTATTAATGANC 1498
474 IucIleAspHisGlnLysArgSerGluLeuValSerGlnAlaSerLeuTyr 490
::|:||||| ::|:||||| ::|:|||||
1499 AGTATGAGACCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1548
491 LeuGlu.....Th 493
|||:||||| ::|:||||| ::|:|||||
1549 TTTGAAAAAAGAAATGATTAATTAATTAATTAATTAATTAATTAAT 1598
493 rPheHisIleIleGluProIleTyrHisAspAlaPheGlnPhe..... 507
|||:||||| ::|:||||| ::|:|||||
1599 CAATCATCTATTTCAGAA.....AATGATGCTTTCAATTTGAAGACTA 1642

```

508AlaMetAsnLysLysLeuSerAsnLeuGlyVal 518
 1643 CTATATTTTAAATCCGGCAGAGATGAATATTAAACCAAGAGAGATT 1692
 519 SerProthrgly 522
 1693 AATTGTGTAGGC 1704

seq_name: /c9n2_6/plodata/2/1na/6B_COMB.seq:US-08-462-467B-3

seq_documentation block:

Sequence 3, Application US/08462467B
 Patent No. 6210899
 GENERAL INFORMATION:
 APPLICANT: Rosenbaum, Jan S
 TITLE OF INVENTION: The use of a BMP Protein Receptor
 TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
 TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
 NUMBER OF SEQUENCES: 39
 CORRESPONDENCE ADDRESS:
 ADDRESS: The Procter & Gamble Company
 STREET: 11810 East Miami River Road
 CITY: Ross
 STATE: OH
 COUNTRY: USA
 ZIP: 45061
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/462,467B
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Herstro, Bart S
 REGISTRATION NUMBER: 32,572
 REFERENCE/DOCKET NUMBER: 5474R
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (513) 627-0633
 TELEFAX: (513) 627-0260
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2156 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-462-467B-3

alignment_scores:

Quality: 117.00 Length: 485
 Ratio: 0.544 Caps: 24
 Percent Similarity: 44.330 Percent Identity: 21.031

alignment block:

US-09-824-567-2 x US-08-462-467B-3/rev ..

Align seq 1/1 to reverse of: US-08-462-467B-3 from: 1 to: 2156

4 IISerValGlyIleCysIleThrIle.....LeuLeu 15
 1982 ATAGCGATTGACATTGATGCTTGGGCTCACAGATTGTTCTTTC 1933
 15 IISerValValLeuGlnGlyCysLysSerIleSerIleThr 32
 1932 CCAATCATCATTAAGTTCAGCAATCCCTTCTCAGACACAGCTGCAATA 1983
 32 eIrargLylGluLeuAlaIleAsnIleArgAspGluProArgSerLeuasp 48

1882 GCGGAGCCCTGCATCCTGGTCCCAACAGCTTCGATTGCTCCTGAG 1833
 49 ProArgGlnValArgLeuLeuSerGlnIleSerLeuValIlyshIleTy 65
 1832 GACCTCACTGCCAGGCTATTCT..... 1809
 65 rGlnGlyLeuValGlnGlnAsnLeuSerGlyAsnIleGluProAla 82
 1808TTCAGGCTCTGCAACTTG..... 1788
 82 eAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLys 98
 1787GCTCTGTTTTCCTGAGAC 1767
 99 LeuLysSerAlaPhe.....TrpSerAsnGlyAspProLeuThr 112
 1766 ACGAAGACCTGATTCCTCAAAAGGAGATGTTCCCAACCTCTGCTG 1717
 112 IagLAspPheIleGluSerTrpGlnValAlaThrGlnGluValSer 128
 1716 AAAG.....CMTCTGACTCTGTGCGGATTCCTCCGGAAGA 1676
 129 GlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleG 145
 1675 GCTCTGATCATCATTAATATCTCC.....ATAGATTAGTCA 1635
 145 nGluGlyHisLeuSerIleAspHisPheGlyValHisProAsnGln 162
 1634 AAGGATCATCTACTGTTTCAAAAGGTGATTCACAGCTCCCTCAAGT 1585
 162 eThrLeuValValThrLeuGluSerPro..... 171
 1584 CACAGCTCTCTGACGACTCTGTGCGGATTCATCTGATGCTGCACT 1535
 172ThrSerHisPheLeuLysLeuAlaLeuProValPhePhe 186
 1534 CCGTTATGCGTCAATATCTCTCCCTGCGGACCACTATATTC.. 1487
 186 oValHisLysSerClnArgThrLeuGlnSerLysSerLeuProIleAla 203
 1486CAGTCAGCTCATGCAAGCTCAAA 1462
 203 eTyrAsn.....GlyAlaPheTyrPro 208
 1461 GTCCTAATAACAGATTCATCATTTTTCATGAGACATTTCTGCTGT 1412
 209 LysAsnIleLysGlnLysGlnIlePheIleLysLeuSerLysAsnProHis 225
 1411 TTAAATCTCATGGAAATTGCAAGTT..... 1385
 225 TTYrAsn..... 227
 1384 TATAATGATCTCTGCTGTAATCTGTGAGAGATAACCCAGCTCT 1336
 228GlnSerGlnValGlnThrLysThrIleThrIleHisPheIlePro 242
 1335 AGTAACAGAAATGAGCAAGCGCAAGACTTACCACTGCTGTGGA 1286
 243 AspAlaAsnThrAlaIleLysLeuPheAsnGlnGlyLysLeuAsnTrpG 259
 1285 GACTTAATACTTGCA.....AAGATCATTTGGA 1254
 259 nGlyProProTrpGlyGluArgIlePro.....GlnGluThrLys 273
 1253 TAGNACTCATTCACAGCAAAATATTCATGGTCCATCTGCACTGACT 1204
 273 eIraSnLeuGln..SerLysGlyHisLeuHisSerPheAspValAlaGly 289
 1203 CTCATCTCAACTATTAACGGGCAATGTTGATCTGATTCATCAAAAGCA 1154
 289 hrSerTrpLeuThrPheAsnIle..... 296

```

1 MOLECULE TYPE: CDNA
2 FEATURE:
3 NAME/KEY: CDS
4 LOCATION: join(409..2154)
5 US-08-334-179A-3
6
7 Alignment_scores:
8     quality: 117.00      Length: 485
9     Ratio: 0.544        Gaps: 24
10    Percent Similarity: 44.330    Percent Identity: 21.031
11
12 Alignment_block:
13 US-09-824-567-2 x US-08-334-179A-3/rev ..
14
15 Align seg 1/1 to reverse of: US-08-334-179A-3 from: 1 to: 2157
16
17     4  llservValGlyIleCysIleIle.....LeuLense 15
18     |||||:|||||
19 1982 ATGACGATGACATTTGATGATGACGTGGGCTCAGACGATTTGTTCTTC 1933
20
21     15  rLuserValValleuGInGlyCysIlyGuserSerHisserSerThrs 32
22     |.....:
23 1932 CAAATCATCATAAAGTTCAGGACATTCCTTCCATGACACACTGTGAGTAA 1883
24
25     32  eArGgIgluLeuAlaIleasnIleArgAspIuProArgSerLeuasp 48
26     |||||:|||||:.....:
27 1882 GCGAGGCTCTGATCCGTCGGTCCACAGACTTCGATGTCCTTGAGT 1833
28
29     49  PrArGInhValArgLeuLuserGluIleSerLeuValIlyHisIleTy 65
30     |||||:|||||:.....:
31 1832 GACCTCATGCGCAGGCTTTTCT..... 1809
32
33     65  rGIuGIlyLeuValGIInGluAsnAsnLeuSerGIlyAsnIleGIuProAla 82
34     |||||:|||||:.....:
35 1808 .....TTCCAGGCTCTCGGAGACTT..... 1788
36
37     82  euAlaGluSprIySerLeuSerSerSpIyLeuThrIyThrIlyHis 98
38     |||||:|||||:.....:
39 1787 .....GCTCTGTTTCCCTAGAC 1767
40
41     99  leuLysSerAlaIle.....TTPSerAsnGIlyAspProLeuThr 112
42     |||||:|||||:.....:
43 1766 ACGAGACCTGATTCCTCAAAAGTGAGTGTTCACACTCTGCTG 1717
44
45     112  IaGluAspHeIleGluSerTrpIyScInValAlaThGIuGIuValSer 128
46     :|||
47 1716 AAAAG.....CCATCTGTACTCTGTAGGATTCCTCCGTGGAGA 1676
48
49     129  GIlyIleTyAlaPheAlaLeuAsnProIleLysAsnValArgIlyIleGI 145
50     |||||:|||||:.....:
51 1675 GGCTCTACATCTCATTAATATCTCC.....AATGAATTAAGTCA 1635
52
53     145  nGIuGIyHisLeuSerIleAspHisPheGIyValHisSerProAsnIus 162
54     |||||:|||||:.....:
55 1634 AGAGCATACATGCTACTGTTTCAAGCTGATTCACAGTCCCTCAAGTT 1585
56
57     162  eThrIleuValAlaThrIleuGIuserPro..... 171
58     |||||:|||||:.....:
59 1584 CACAGCTCTTTCACACTTCTGGTCCATATATGATTAAGTCCAACT 1535
60
61     172  .....ThSerHisPheLeuLysLeuAlaIleuProValIlePhePr 186
62     |||||:|||||:.....:
63 1534 CGCTTATGAGCTGATTAATCTTCCTCCCGGGGAGCACAGCTATTTTC 1487
64
65     186  oValHisIySerGIlnArgThrIleuGIuserIlySerLeuProIleAlas 203
66     |||||:|||||:.....:
67 1486 .....CAGTCACCTTCATGACAGCACTCCAA 1462
68
69     203  eT.....GIlyAlaPheTyPro 208
70     |||||:|||||:.....:
71 1461 GTCACTAATACAGAGTTCATCATTTTTCATGTAGGACATTTCTGCTGT 1412

```

```

1  CITY: ROSS
2  STATE: OH
3  COUNTRY: US
4  ZIP: 45061
5
6  COMPUTER READABLE FORM:
7
8  MEDIUM TYPE: Floppy disk
9  COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: Patent In Release #1.30, Version #1.30
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/34,179A
16
17 FILING DATE: 04-NOV-1994
18
19 CLASSIFICATION: 435
20
21 ATTORNEY/AGENT INFORMATION:
22
23 NAME: CONSTANJE, BRAH J.
24 REGISTRATION NUMBER: 34,804
25 REFERENCE/DOCKET NUMBER: 3473
26
27 TELECOMMUNICATION INFORMATION:
28
29 TELEPHONE: 513-627-2858
30 TELEFAX: 513-627-0260
31
32 INFORMATION FOR SEQ ID NO: 1:
33
34 SEQUENCE CHARACTERISTICS:
35
36 LENGTH: 3601 base pairs
37
38 TYPE: nucleic acid
39
40 STRANDEDNESS: double
41
42 TOPOLOGY: linear
43
44 MOLECULE TYPE: cDNA
45
46 FEATURE:
47
48 NAME/KEY: CDS
49
50 LOCATION: join(409..3522)
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986

```

Align seg 1/1 to reverse of: US-08-334-179A-1 from: 1 to: 3601

```

4 11eserValgIleCySierHire.....Leuense 15
   |||||  || |||||
1982 ATAGCAGTATGACATTGATGATGATGCTGCACACATTGTTCTTTC 1933
15 rLeuserValValIleuInclYcysIysoluserSerHisSerThrS 32
   : : : : : : : : : : : : : : : : : : : : : :
1932 CCAAATCATCATATGATTCACCCATCCTTCTCAGCAGCACTGTGCGATGA 1883
32 eArArgGlyGluLeuAlaIleasnIleArgaspGluProArgSerLeuasp 48
   |||||  |||||  : : : : : : : : : : : : : : : :
1882 GCGAGCCTCTGCATCCTGCTGGTCCCAACAGTCTTGATGTTCTCCTGAGT 1833
49 ProArgInValnArgLeuSerGluIleSerLeuValLysHisIleTy 65
   : : : : : : : : : : : : : : : : : : : : : :
1832 GACCTCACTGCGCAGGCTATTCT..... 1809
65 rGluGlyLeuValGluIasnAsnLeuSerGlyAsnIleGluProAlaI 82
   : : : : : : : : : : : : : : : : : : : : : :
1808 .....TTCCAGGCTTCGGACCTG..... 1788
82 euAlaGluAspTyrSerLeuSerSerAspGlyLeuThrIleThrPheLys 98
1787 .....GCTCTGCTTTTCCCTAAC 1767
99 LeuYsserAlaPhe.....TyrSerAsnGlyAspProLeuThra 112
   : : : : : : : : : : : : : : : : : : : : : :
1766 ACGAGACCTGATATCCCAAAAGTGGAGGTTCACACTCTGTCTGTG 1717
112 LagIuAspPheIleGluSerTrpLysGluValAlaIleThrGlnValSer 128
   : : : : : : : : : : : : : : : : : : : : : :

```


1716 AAAAG.....CCATCTGCTACTCTGTGACGATTCCTCCGAGACA 1676
129 GYIIEYRALPheAlaLeuAsnProIlElysAsnValArgLysIleG1 145
1675 GGTCTGACATTCATTAATATCTCC.....AATGATTAATGTCOA 1635
145 nG1UG1YHISLeuSerIleAspHisPheGlyValHisSerProAsnG1us 162
1634 AGACATATCATCTGTCTGTTTCAAGCTGATTCACAGTCCCTCAAGTT 1585
162 erThrLeuValThrLeuGlnSerPro..... 171
1584 CACAGCTCTCTTACGACTCTCGGCGCATATATCTGATAGGCCAACCT 1535
172ThSerHisPheLeuLysLeuAlaLeuProValPhePhePr 186
1534 CGCTTATGCTGCTATATCTCTCCCTCGCGCACAGCTGATTTCT.. 1467
186 oValHisLysSerGlnArgThrLeuGlnSerLysSerLeuProIleAla 203
146CAGTCAGCTCATGACAGCCCAA 1462
203 er.....GlyValPheThrPro 208
1461 GTCATAATAACACAGTTCATATTTTCATAGACATTTCTGCTGT 1412
209 LysAsnIleLysGlnLysGlnThrPheLysLeuSerLysAsnProHisL 225
1411 TTTAAATCTCGATGGAAATTCAGATT..... 185
225 TTYAsn..... 227
1384 .TATAATGATCTCCCTGCTGATTCCTGCTGATTAAGCAGCTCTCT 1336
228GlnSerGlnValGlnThrLysThrIleThrIleHisPheIlePro 242
1335 AGTAACAGATAGACAGACGCAAGAGCTTACCCAGTCTGTGTGGA 1286
243 AspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpG1 259
1285 GACTTAATACTTCGATA.....AAGATCCATTGGGA 1254
259 nGlyProProThrPheLysLysArgLeuPro.....GlnGlnThrLeuS 273
1253 TAGTACTTCATCACAACAAATATTCATGCTGCTGATTCGACGACTCT 1204
273 erAsnLeuGln..SerLysGlnHisLeuHisSerPheAspValAlaGlyT 289
1203 CTCATCTCCAACTATAAGCGGCAATGTTTCATGTTCCATCAAGGCA 1154
289 hISerTrpLeuThrPheAsnIle..... 296
1153 CTCTGTAATAGTCTTCTGCTGATTAATTCATGACGTTTGCAGAAAGGA 1104
297 AsnLysPheProLeuAsnAsnMetLysLeuArgLysAlaLeuAlaSerAl 313
1103 AACACTTTTACAGCACTGACGCTCATCCAAAGAGCCTTATATATCTCT 1054
313 AlEusAspLysGlnAlaLeuValSer.....ThrIlePheLeuGlyAla 328
1053 TCCATATTCGACCTCGGCAATGCTCCACAGCTTTCAGATTATCTAGAT 1004
328 lAlLysThrAlaAspHisLeuLeu..... 335
1003 CAAAGAGGCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 954
336ProThrAsnIleHisSerLysTrpGlnHisGln.. 346
953 CTTTGTATACGCTCTCTCTCAACATP...CTGATCCAAACATAAGCC 907
347LysGlnGlnMetAlaGlnArgGlnAlaLysLysLysLeuP 361
906 AACTATCAAAACAGCTAATACAGAGACTGATGCAAAAGCAATGATTATTG 857

361 heuysGlnAlaLeuGlnGlu.....LeuGlnIleThrAlaLysAs 374
856 TCTCATCTGCTGTTAATGAATGAGTGAGTGAAGTGCTGTTGTCAGA 807
374 pleuGlu.....HisLeuAsnLeuIlePhePro 384
806 GGTGAAATTCCTCAATGAAGTTGACATTACATTAATCTCTGCTAC... 761
384 alSerSerAlaSerSerLeuLeuValGlnLeuIleArgGlnGlnTrp 400
760ACAGAGAAAGGATGATTCCTCATCTGATTCGAGGAGAGTGG 716
seq_name: /cgn2_6/ptodata/2/ina/6b.COMB.seq:us-08-462-467B-1
seq_documentation_block:
; Sequence 1, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road
; CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hersko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3603 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-462-467B-1
alignment_scores:
Quality: 117.00 Length: 485
Ratio: 0.544 Gaps: 24
Percent Similarity: 44.330 Percent Identity: 21.031
alignment_block:
US-09-824-567-2 x US-08-462-467B-1/rev ..
Align seg 1/1 to reverse of: US-08-462-467B-1 from: 1 to: 3603
4 lIeSerValGlyIleCysIleThrIle.....LeuLeuS 15
1982 ATAGCAGTAAGACATGATGATGCTGTGAGCTCACACATTTGTTCTTTC 1933
15 lIeuserValValLeuGlnGlyCysLysLysSerHisSerHis 32

NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Klebsiella pneumoniae
US-08-743-637B-8

Alignment_scores:
Quality: 113.50 Length: 87
Ratio: 2.027 Gaps: 3
Percent Similarity: 64.368 Percent Identity: 34.483

Alignment_block:
US-09-824-567-2 x US-08-743-637B-8/rev ..

Align seq 1/1 to reverse of: US-08-743-637B-8 from: 1 to: 238

```
92 GlyLeuThrTyrThrPhelysLeuLysSerAlaPheTrpSerAsnGlyAs 108
|||||
238 GGCTGACC.....TGCTGACGGGCAC 216

108 PProLeuThrAlaGluAspPheIleGluSerTrpLysGlnValAlaThrG 125
|||||
215 GGCCATTACAGCGCAGATATGCTGTGAGCTGGCAACGGCTGCTCGC 166

125 InGluValSerGlyIleTyrAla...PheAlaLeuAsn...ProIleLys 139
|||||
165 CGGCACACACCTCCCTACGCCAGCTACCTGGCAATATGCATATCGCC 116

140 AsnValArgLysIleGlnGluHisLeuSerIleAspHisPheGlyVal 156
|||||
115 AACGCCCGGAGATTGCCCTCGGGCAAAAGGAGACCGGAGAGCTGGGGGT 66

156 HisSerProAsnGluSerThrLeuValAlaThrLeuGluSerProThr 173
|||
65 GAAAGCGCTGAACGACACACCGCTGCAGGTCACTGTGACCCAGCGAATG 16

173 eHisPheLeu 176
:: |||||
15 CCGCCTCTCTG 5
```

seq_name: /cgn2_6/prodata/2/ina/6A_COMB.seq:US-08-526-840B-8

seq_documentation_block:
Sequence 8, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Klebsiella pneumoniae
US-08-526-840B-8

Alignment_scores:
Quality: 113.50 Length: 87
Ratio: 2.027 Gaps: 3
Percent Similarity: 64.368 Percent Identity: 34.483

Alignment_block:
US-09-824-567-2 x US-08-526-840B-8/rev ..

Align seq 1/1 to reverse of: US-08-526-840B-8 from: 1 to: 238

```
92 GlyLeuThrTyrThrPhelysLeuLysSerAlaPheTrpSerAsnGlyAs 108
|||||
238 GGCTGACC.....TGCTGACGGGCAC 216

108 PProLeuThrAlaGluAspPheIleGluSerTrpLysGlnValAlaThrG 125
|||||
215 GGCCATTACAGCGCAGATATGCTGTGAGCTGGCAACGGCTGCTCGC 166

125 InGluValSerGlyIleTyrAla...PheAlaLeuAsn...ProIleLys 139
|||||
165 CGGCACACACCTCCCTACGCCAGCTACCTGGCAATATGCATATCGCC 116

140 AsnValArgLysIleGlnGluHisLeuSerIleAspHisPheGlyVal 156
|||||
115 AACGCCCGGAGATTGCCCTCGGGCAAAAGGAGACCGGAGAGCTGGGGGT 66

156 HisSerProAsnGluSerThrLeuValAlaThrLeuGluSerProThr 173
|||
65 GAAAGCGCTGAACGACACACCGCTGCAGGTCACTGTGACCCAGCGAATG 16

173 eHisPheLeu 176
:: |||||
15 CCGCCTCTCTG 5
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 04:38:12 ; Search time 37.11 Seconds
(without alignments)
555,074 Million cell updates/sec

Title: US-09-824-567-2
Perfect score: 2739
Sequence: 1 MKRISVGICITILLISVTL.....LSNCGSPGVDFRIAKEN 532

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	DB ID	Description
1	614.5	22.4	545	1	OPPA_BACSU
2	489.5	17.9	543	1	DPEE_BACSU
3	438.5	16.0	537	1	MPPA_ECOLI
4	415	15.2	543	1	OPPA_ECOLI
5	410	15.0	543	1	OPPA_SALT
6	397.5	14.5	535	1	YGIS_ECOLI
7	370	13.5	541	1	OPPA_HAEN
8	354.5	12.9	514	1	Y213_HAEN
9	337	12.3	652	1	ALIB_STRPN
10	301	11.0	524	1	NKKA_ECOLI
11	291.5	10.6	547	1	HBPA_HAEN
12	287.5	10.5	516	1	YDGS_ECOLI
13	283.5	10.4	535	1	DPPA_ECOLI
14	275.5	10.1	659	1	AMIA_STRPN
15	272.5	9.9	660	1	ALIA_STRPN
16	249	9.1	512	1	YLIB_ECOLI
17	245	8.9	549	1	SAPA_SALT
18	242.5	8.9	543	1	APPA_BACSU
19	238	8.7	565	1	SAPA_HAEN
20	212	7.7	547	1	SAPA_ECOLI
21	191.5	6.6	600	1	OPPA_LACI
22	165.5	6.0	600	1	OPPA_LACI
23	164	6.0	542	1	XP55_STRPI
24	164	6.0	591	1	YC80_MYCTU
25	156.5	5.7	693	1	AGPA_RHIME
26	153.5	5.6	531	1	Y470_RHISN
27	148	5.4	663	1	Y4WM_RHISN
28	143	5.2	604	1	YELA_ECOLI
29	141.5	5.2	744	1	SKX1_HAEN
30	126.5	4.6	1432	1	SK13_YEAST
31	120.5	4.4	3726	1	ABPI_MOUSE
32	114.5	4.2	412	1	HOFO_ECOLI
33	111	4.1	566	1	YBAE_ECOLI

34	110	4.0	1807	1	VT42_XENLA	P18709 xenopus lae
35	109	4.0	441	1	HS82_ASPFU	P40292 aspergillus
36	108.5	4.0	1237	1	YDY2_SCHPO	O13683 schizosach
37	107	3.9	623	1	RECG_HELPU	O92181 helicobacte
38	107	3.9	701	1	HS90_PODAN	O43109 podospora a
39	107	3.9	1089	1	NMD2_YEAST	P38798 saccharomyc
40	106.5	3.9	2607	1	BACB_BACLI	O68007 bacillus li
41	106	3.9	1161	1	FOL_SFY1	P23074 simian foam
42	104.5	3.8	557	1	INRI_HUMAN	P17181 homo sapien
43	103.5	3.8	4639	1	DYHC_DROME	P37276 drosophilla
44	103	3.8	436	1	ZPI_RAT	O62975 rattus norv
45	103	3.8	580	1	MUTL_CHLPPN	O92794 chlamydia p

ALIGNMENTS

```

RESULT 1
ID OPPA_BACSU STANDARD: PRT; 545 AA.
AC P24141; P23399;
DE 01-MAR-1992 (rel. 21, Created)
DE 01-MAR-1992 (rel. 21, Last sequence update)
DE 16-OCT-2001 (rel. 40, Last annotation update)
DE Oligopeptide-binding protein oppa precursor.
GN OPPA OR SPOOKA.
OS Bacillus subtilis.
OC Bacteria: Firmicutes: Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
CX NBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91194547; PubMed=1901616;
RA Perego M., Higgins C.F., Pearce S.R., Gallagher M.P., Hoch J.A.;
RT "The oligopeptide transport system of Bacillus subtilis plays a role
RL in the initiation of sporulation."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91139580; PubMed=1899858;
RA Rudner D.Z., Ledoux J.R., Ireton K., Grossman A.D.;
RT "The spo0K locus of Bacillus subtilis is homologous to the
RT oligopeptide permease locus and is required for sporulation and
RT competence."
RN [3]
RP Bacteriol. 173:1388-1398(1991).
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE OLIGOPEPTIDE
CC PERMEASE, A BINDING PROTEIN-DEPENDENT TRANSPORT SYSTEM, IT BINDS
CC PEPTIDES UP TO FIVE AMINO ACIDS LONG WITH HIGH AFFINITY. ALSO
CC REQUIRED FOR SPOULATION AND COMPETENCE.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X56347; CA39787.1; -
CC EMBL: M57689; AAA62687.1; -
CC EMBL: Z99110; CAB1500.1; -
CC PIR: S15230; S15230.
CC PIR: A38447; A38447.
CC HS5P: P06202; 1852.
CC Subtilist; BG10771; oppa.
CC Interpro: IPR000914; SBP_bac_5.
CC Pfam: PF00496; SBP_bac_5; 1.

```

DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
 KW Peptide transport; Transport; Membrane; Signal; Sporulation;
 Lipoprotein; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 545
 FT LIPID 21 21
 FT LIPID 21 21
 FT CONFLICT 26 26
 FT CONFLICT 195 195
 FT CONFLICT 340 340
 SQ SEQUENCE 545 AA: 61524 MW: 8846999DC4B4B6 CRC64;

Query Match 22.4%; Score 614.5; DB 1; Length 545;
 Best Local Similarity 29.3%; Pred. No. 8e-36;
 Matches 158; Conservative 108; Mismatches 249; Indels 25; Gaps 13;

QY 10 ITLLSLVYLOGC-----KSSHSSTSGEALNIRDEPNSLDPROVRLSEISLV 61
 DB 8 VTMLLFTLVLSACGFGGTGSGNGKSKGKKTJTNINIKTEPFLHPGLANDSVSGVY 67
 QY 62 KHIEGLVQENNLGNIEPALAEDYSLSDGLTYFYLKSAF-MSNGDPLADPTEISWK 120
 DB 68 RQFEELTR-IMADSEPEEGMAKIEISKDGKTYFTTRIDGKKSNDPVTADPFEYAWK 126
 QY 121 QVATQVSGITVAFALNPYKYNKIQEGHLSIDHFGVSPNESTLVLTLESPTSHFLKLLA 180
 DB 127 WALDPNNSQVAYQLYYIKGAEMANTGKSLDVAKAVNDKTLKVELNPTPYETELTA 186
 QY 181 LPVFPVHKSORTLOSKSLPIASGAFPKY--IKOKW-----IKLSKNHYVNOQVET 233
 DB 187 FYTYMPLNE-KIAEKNNKMWNTNADDDVNSNGPKMTAMKHSGSTLTKNQYWDKDKVXL 245
 QY 234 KTTTHIFPDANTAAKLFENQCKLWQSPWGERIQEFTLSNOSKGLHSEFVAGTSWLT 293
 DB 246 KXIDVYMINNNNTTELKFGQACELDMACMLPGQ-LPTESLPLTKDGLHNPYLAGYVYWK 304
 QY 294 ENINKEPLNNMKLRALASALDKALYSTIFIGRAKTAHDHLLPTNIHSYEHOKQEMAOR 353
 DB 305 FNTAEKPLDNNVIRKALYSLDROSTYKNTQGEQPMMAVPEPTMGPFNDKEGYFKN 364
 QY 354 QA-YAKLFEKALELEIQT-AKDLHNLIFPVSSASSLLVOLIROMKESIGFAIPYV 411
 DB 365 DVTAKETLEKTKEMGSKASDLPKIKLSYN-TDDAIAKIAQAVQEMKKNLGVDELD 423
 QY 412 GKFEALLQADSSGNSFSLATGWFADPMAFLITFVYPS-GVPPYAINHKDFLEILON 470
 DB 424 NSEMNYYIDKLHSQDYQIGRMGLDGNFINLELFRDXNGANNNGWENPEFKTLNQ 483
 QY 471 IEEDOHKRSRLVSA-QA-SLYLEFTHIEPTIHDAFOFAMANKKLSNLGVSPTGVDFRYA 529
 DB 484 SQTEIDKTRKRLKKAAGIFIDEMP-VAPITYTDTWVDENLKVIMEGTGEVYFRNA 542

RESULT 2
 DPPE_BACSU STANDARD; PRT; 543 AA.
 AC P26906; O34801;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dipeptide-binding protein dppe precursor.
 GN DPPE OR DCIAR
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=92114768; PubMed=1766370;
 RA Mathopoulos C., Mueller J.P., Slack F.J., Murphy C.G., Patankar S.,
 Bukusoglu G., Sonenshein A.L.;

RT "A Bacillus subtilis dipeptide transport system expressed early
 RT during sporulation."
 RL Mol. Microbiol. 5:1903-1913(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RL "Sequence of the Bacillus subtilis genome between xlyA and ykoR";
 Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC FOR DIPEPTIDES. PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC WITH HIGH AFFINITY. IS EXPRESSED TO FACILITATE ADAPTATION TO
 CC NUTRIENT DEFICIENCY CONDITIONS, WHICH ALSO INDUCE SPOULATION.
 CC -1 SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (probable)
 CC -1 DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING SPOULATION.
 CC -1 INDUCTION: NUTRIENT DEFICIENCY CONDITIONS.
 CC -1 SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 CC PROTEIN FAMILY 5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56678; CAA0006.1; -
 DR EMBL; AJ002571; CAA05576.1; ALT_INIT.
 DR EMBL; Z99110; CAB1315.1; ALT_INIT.
 DR PIR; S16651; S16651.
 DR HSSP; P06202; 1B52.
 DR Subtilist; BG10846; dppe.
 DR InterPro; IPR000914; SBP_bac-5.
 DR Pfam; PF00496; SBP_bac-5; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
 KW Peptide transport; Transport; Membrane; Signal; Sporulation;
 KW Lipoprotein; Complete proteome.
 FT SIGNAL 1 22
 FT CHAIN 23 543
 FT LIPID 23 23
 FT CONFLICT 339 339
 SQ SEQUENCE 543 AA: 61822 MW: 723434619413356 CRC64;

Query Match 17.9%; Score 489.5; DB 1; Length 543;
 Best Local Similarity 26.0%; Pred. No. 4.9e-27;
 Matches 143; Conservative 116; Mismatches 244; Indels 47; Gaps 13;

QY 12 ITLLSLVYLOGC-----KSSHSSTSGEALNIRDEPNSLDPROVRLSEISLV 60
 DB 12 LALGLSALMGCTANQAKGKESGSHKAKTSGEYVLYVNNNEPSTFDPDLC--FNNNSWQ 69
 QY 61 -VKHIEGLVQENNLGNIEPALAEDYSLSDGLTYFYLK-SAFWSNGDPLTADPFI 118
 DB 70 PLANNIMEGLRLGK-DHEPEPAMAEKWSYKDKKTYFTIRNAKNTMGDVTAGDEYFA 128
 QY 119 WKQVATQVSGITVAFALNPYKYNKIQEGHLSIDHFGVSPNESTLVLTLESPTSHFLK 178
 DB 129 WKRMPLDKKGAASSAFIIFYIEGGEAYNSGKGDVYKATKADRTLEATLAPQKFLSV 188
 QY 179 LALPVEFPVHK-----SORTLOSKSLPIASGAFPKYKINIKQKWLKLSKNHYVNOQVET 233
 DB 189 VSNDAIFPVNEVYKONKMPAESDITVNGPDKLTETKMHDSITMKSQDYMDKDTVXL 248
 QY 234 KTTTHIFPDANTAAKLFENQCKLWQSPWGERIQEFTLSNOSKGLHSEFVAGTSWLT 293
 DB 249 DKYKMWAVSDRNTDYQMGQSELD-----TAVPAEISDQLLDQDNNTVDAAGILYFPR 302
 QY 294 ENINKEPLNNMKLRALASALDKALYSTIFIGRAKTAHDHLLPTNIHSYEHOKQEMA-- 351

Db 303 FVNMPEPQENIRKAFAMAVDOEEIVKYVTKNNKETA-HAFVSPGTPQDCKDREAGG 361
 QY 352 ----QROAVAKKLEKKEALELOITAKDEHLNLFPPVSSASR-----LLVOLIREQKE 402
 Db 362 DLIKNESKAKOLEKGMKE-----ENVNKKLPATLLIYSPKPEKKIABAIOQKILKN 413
 QY 403 SLGFAIPVIGKEFALLQADLSGNSFLATGCGWFDPPADPMAPFLITAIYSSGPVPAIHK 462
 Db 414 SLGVGVYKLANMENVPLEDQKALKRQFQSSSFLPDYADPISFLEAFQTSNKNRTGMANK 473
 QY 463 DPLELILQNIQEDODHQRSELSVQSALYLETFPHIIEFYHDAFQFAMKRLSLNLCVSPFG 522
 Db 474 EYDQILQKAKNEADEKTRFSIMHQAEELLINAPILIVFYVYQVHILQNVQVIGVIRHVG 533
 QY 523 VVDFYAKEN 532
 Db 534 YIDLKWKADKN 543

RESULT 3
 ID MPPA_ECOLI STANDARD; PRT; 537 AA.
 AC P77348; 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Periplasmic murein peptide-binding protein precursor.
 GN MPPA OR B1329.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid:562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-32.
 RC STRAIN-K12 / AT980;
 RX MEDLINE=98155149; PubMed=9495761;
 RA Pak J.T., Raychaudhuri D., Li H., Normark S., Mengin-Lecreulx D.;
 RT "Mppa, a periplasmic binding protein essential for import of the
 bacterial cell wall peptide L-alanyl-gamma-D-glutamyl-meso-
 diaminopimelate."
 RL J. Bacteriol. 180:1215-1223(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426517; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakase S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sampei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28,040,1 min region on the linkage map."
 RL DNA Res. 3:363-377(1996).
 CC -1- FUNCTION: ESSENTIAL FOR THE UPTAKE OF THE MUREIN PEPTIDE L-ALANYL-
 GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE. ALSO TRANSPORTS SOME ALPHA-
 LINKED PEPTIDES SUCH AS PRO-PHE-LYS WITH LOW AFFINITY. THE
 TRANSPORT IS EFFECTED BY THE OLIGOPEPTIDE PERMEASE SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; U88242; AAC38216.1; ALT_INIT.
 DR EMBL; AE000231; AAC74411.1; ALT_INIT.
 DR EMBL; D90772; BAAL1932.1; ALT_INIT.
 DR EMBL; D90771; BAAL1922.1; ALT_INIT.
 DR HSSP; P06202; IJEV.
 DR SWISS-2DPAGE; P77348; COLI.
 DR EcoGene; EG13376; mppa.
 DR InterPro; IPR000914; SBP_bac.5.
 DR Pfam; PF00496; SBP_bac.5; 1.
 DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
 KW Peptide transport; Transport; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 22
 FT CHAIN 23 537
 FT PROTEIN PERIPLASMIC MUREIN PEPTIDE-BINDING
 SQ SEQUENCE 537 AA; 59900 MW; C6A17656836DC3AC CRC64;

Query Match 16.0%; Score 438.5; DB 1; Length 537;
 Best Local Similarity 26.0%; Pred. No. 1.9e-23;
 Matches 134; Conservative 100; Mismatches 248; Indels 33; Gaps 12;

QY 4 ISVGCITILLISLVYLOGCKSSHSSTSGELAINIRDEPRSLDPROYRLSEISLYKH 63
 Db 5 VSTTCALIVSSISLSYAAEVPSGVTLAEKQELVHRIKDEPASLDPKAVGLPELQVIRD 64
 QY 64 IYEGVJOENNNISGNIPALAEYDLSLSDGLYTFKRL-SAFWSNGDPLAEDFISMKOV 122
 Db 65 LEFGIVKQNE-KGEIVDGYATQWK-SNDNRIMTFTLBDNKKAKMDGTPVYADQFVYSWQRL 122
 QY 123 ATQEVSGTYAE--ALNPINKVKIQEGHLSIDHGVSPNESTLYVTLSPSSHKLKLA 180
 Db 123 VDPKTLSPFAFMAFALAGINNAQAIIIDKATPDQGLVAVNAHLKIQDLKPLPKNVILTA 182
 QY 181 LPVFPEYHK-----SQTLSKSLPTASGATYPPKNIKQKQIKSNPHAYNOSQYERTKI 236
 Db 183 NFAPFPQKANVESGKEMTKPGLNLGNGAVYKERYVNEKLVVYPVTHYMDNKAIVLQKV 242
 QY 237 TIHFIPDNTAAKLFNQGKLNMGCPWGERIPDETLSNLSQK-----GHLSHSDVAGTS 290
 Db 243 TELPINESAKTKRYLAGIDID-----TTSEFPKNMYOKLKLDPQGVYTPPOLGTY 293
 QY 291 WLTNINIKKPPLANMKLREALASALDKALVSTIFLGRAKTAHLLPTNINHSY-PEHOK-Q 348
 Db 294 YYAENTQKGPPTADQVRRLASWTLIDRLMTEKVLGGEKPAHMFPPDVTAGFTPEPPEFE 353
 QY 349 EMAQRO--AVAKKLEFEALEEELQITAKDEHLNLFPPVSSASSLLVOLIREOKESLGF 406
 Db 354 QMSQEEILMAQKILLISAAGYQPKL---LTLVNTSENQKTAI-AVASMKKNLGL 408
 QY 407 AIPVIGKEFALLQADLSGNSFLATGCGWFDADPMAPFLITAIYSSGPVPAIHKDFLE 466
 Db 409 DVKLQNEKMTYIDSNNGNNDVIRASWGDVNEPSTFILLTSHSGNISRENNPAYDK 468
 QY 467 ILQNIQEDODHQRSELSVQSALYLETFPHIIEFY 501
 Db 469 VLAQASTENTVAKRADYNAEKILMEQAPIARIY 503

RESULT 4
 ID MPPA_ECOLI STANDARD; PRT; 543 AA.
 AC P23843; P76829;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Periplasmic oligopeptide-binding protein precursor.

GN OPPA OR B1243.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=90256748; PubMed=2187863;
 RA Kashiwagi K., Yamaguchi Y., Sakai Y., Kobayashi H., Igarashi K.;
 RT "Identification of the polyamine-induced protein as a periplasmic
 RL oligopeptide binding protein.";
 RL J. Biol. Chem. 265:8387-8391(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA Pahel G., Short S.A.;
 RN Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RC MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Siao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RC MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto Y., Nishio Y.,
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horinuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RL corresponding to the 28.0-40.1 mta region on the linkage map.";
 RN DNA Res. 3:363-377(1996).
 RN [5]
 RP SEQUENCE OF 1-53 FROM N.A.
 RC STRAIN-K12;
 RC MEDLINE=91200315; PubMed=2015910;
 RA Kessler D., Leibrecht I., Knapp J.;
 RT "Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase
 RL activities of Escherichia coli reside on a polymeric protein particle
 RL encoded by adhP.";
 RL FEBS Lett. 281:59-63(1991).
 RN [6]
 RP SEQUENCE OF 1-29 FROM N.A.
 RC STRAIN-DR112;
 RA Igarashi K.;
 RN Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 27-37.
 RC STRAIN-K12 / W3110;
 RA Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
 RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
 RA Hochstrasser D.F.;
 RN Submitted (SEP-1994) to the SWISS-PROT data bank.
 RN [8]
 RP SEQUENCE OF 27-38.
 RC STRAIN-K12 / EWG2;
 RC MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RL in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE OLIGOPEPTIDE
 CC PERMEASE, A BINDING PROTEIN-DEPENDENT TRANSPORT SYSTEM, IT BINDS
 CC PEPTIDES UP TO FIVE AMINO ACIDS LONG WITH HIGH AFFINITY.

CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 CC PROTEIN FAMILY 5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, J05433; AAA21302.1; -;
 DR EMBL, M60918; AAB00918.1; -;
 DR EMBL, AE000222; AAC74325.1; -;
 DR EMBL, D90852; BAA16038.1; -;
 DR EMBL, D90763; BAA14775.1; -;
 DR EMBL, X59501; CAA42089.1; -;
 DR EMBL, D83137; BAA11814.1; -;
 DR PIR, A36263; A36263.
 DR HSP, P06202; 1REV.
 DR SWISS-2DPAGE; P23843; COLI.
 DR EcoGene; E810674; oppa.
 DR InterPro; IPR00914; SPP_bac-5.
 DR Pfam; PF00496; SPP_bac-5; 1.
 DR PROSITE; PS01040; SPP_BACTERIAL_5; 1.
 KW Peptide transport; Transport; Periplasmic; Signal; Complete proteome.
 FT STGNAL 1 26
 FT CHAIN 27 543
 FT DISULFID 297 443
 FT CONFLICT 271 271 N -> Y (IN REF. 2).
 FT CONFLICT 314 315 RV -> LW (IN REF. 2).
 FT CONFLICT 487 488 OR -> HG (IN REF. 2).
 SQ SQ000000 543 AA; 60899 MM; BBFF9FBD042254EF CRC04;
 Query Match 15.2%; Score 415; DB 1; Length 543;
 Best local similarity 25.3%; Pred. No. 8.5e-22;
 Matches 134; Conservative 108; Mismatches 229; Indels 58; Gaps 16;
 QY 4 ISVGICITLLSLSVYLGCKESHSSTSGELAINIDEPRSDPPQVRLSEISLVKH 63
 DB 10 VAAGV-LAALMAGNVALLADVPAGVTLAKQTLVNNSEVSDPHKIGVPESSNSRD 68
 QY 64 IYGVLENNLSGNIEPLAEDYSLSSDGLTYFTPL-KSAFWSGDLTAEEDTESMKRY 122
 DB 69 LPEGLL-VSDLDGAPAGVAVESWD-NKDAKVTFTLRDAAKWSGCTPVTAQDFVSWRS 126
 QY 123 ATOEVSGIYAFAL-NPIKNVKRIQGHSLIDFGVHSPNESTLVYLTSPSHPLKLA 180
 DB 127 VDENTASPYASYLGHTAGIDETLEKKRPTDGLVKAIDHDTLEVTLSPEVPEYKLLV 186
 QY 181 LPTFFPVVHS-----ORTLOSKSLPISAGATFPYNNIKOKMYILSNRPYVQSIVETK 234
 DB 187 HPSTSPVPAAILEKFEGRKWTPGNI-VTNGAVYTLADWVNERIVLEKSPVNNKATVIN 245
 QY 235 TITIHIPDAN-TAAKLFGQKLNMQGPPMERIPQETLSLNLSQK-GHLSFVAGTS 290
 DB 246 QVT--YLPISAVYDVNRYRSGEIDMT--NNSPRTILPKLKEIPDEVHYDYLCTY 299
 QY 291 WLTFNINKEPLNNKLEALASALDEALVSTIFGRAKTADHLPTNHSYPERQKQEM 350
 DB 300 YVEINNOKPPNDVVRFTALKLGMDBDILVNV-----KQGNMPPAYGTPPYDGAKL 353
 QY 351 AQ-----RAYAKKILFKALELEDTAKDEHLNLIFFPS-----SASSLIY 393
 DB 354 TQPEWFGNSQKREKAEKLLAEEA-----GYADKPLTLLYNTSDLHKKLAIASSL-- 407
 QY 394 OLIRQWRESLGFPIPVGKEFALLQADSSGNSFSLATGCFADFPADPAFLITAYPSG 453
 DB 408 -----WKKNIQGVKVLVNEKMTFLDRHQCTEPVRAAGMCADYNPEPSLNTMSSS 461
 QY 454 VPPVAINHKDFLEILQNIROBDHOKRSELVSQASLYLETFHTEPIYH 502

Db 462 MNTAHKSPADSTMAETLKVDEAORTALYTKAEQQLDKSDAIVPVY 510

RESULT 5

OPPA_SALTY STANDARD; PRT; 543 AA.

AC P06202;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Periplasmic oligopeptide-binding protein precursor.

GN OPBA OR STM1746.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OX NCBI_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2;

RX MEDLINE=88011222; PubMed=2821267;

RA Hiles I.D., Gallagher M.P., Jamieson D.J., Higgins C.F.;

RT "Molecular characterization of the oligopeptide permease of

RT Salmonella typhimurium."

RL J. Mol. Biol. 195:125-142(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=86274740; PubMed=3525163;

RA Hiles I.D., Higgins C.F.;

RT "Peptide uptake by Salmonella typhimurium. The periplasmic

RT oligopeptide-binding protein."

RL Eur. J. Biochem. 158:561-567(1986).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SCS1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

RT Lt2."

RL Nature 413:852-856(2001).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).

RX MEDLINE=94261830; PubMed=8202710;

RA Tame J.R.H., Murshudov G.N., Dodson E.J., Neil T.K., Dodson G.G.,

RA Higgins C.F., Wilkinson A.J.;

RT "The structural basis of sequence-independent peptide binding by Oppa

RT protein."

RL Science 264:1578-1581(1994).

RN [5]

RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).

RX MEDLINE=96363676; PubMed=8747465;

RA Tame J.R.H., Dodson E.J., Murshudov G.N., Higgins C.F.,

RA Wilkinson A.J.;

RT "The crystal structures of the oligopeptide-binding protein Oppa

RT complexed with tripeptide and tetrapeptide ligands."

RL Structure 3:1395-1406(1995).

RN [6]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX MEDLINE=99349693; PubMed=10422831;

RA Davies T.G., Hubbard R.E., Tame J.R.H.;

RT "Relating structure to thermodynamics: the crystal structures and

RT binding affinity of eight Oppa-peptide complexes."

RL Protein Sci. 8:1432-1444(1999).

RN [7]

RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).

RX MEDLINE=99370085; PubMed=10438628;

RA Sleight S.H., Seavers P.R., Wilkinson A.J., Ladbury J.E., Tame J.R.H.;

RT "Crystallographic and calorimetric analysis of peptide binding to

RT Oppa protein."

J. Mol. Biol. 291:393-415(1999).

-!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE OLIGOPEPTIDE PERMEASE, A BINDING PROTEIN-DEPENDENT TRANSPORT SYSTEM. IT BINDS PEPTIDES UP TO FIVE AMINO ACIDS LONG WITH HIGH AFFINITY.

-!- SUBCELLULAR LOCATION: Periplasmic.

-!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X04194; CAA27785.1; -

DR EMBL; X05491; CAA29039.1; -

DR PIR; A25011; QREBOA.

DR PDB; 1OLA; 31-JUL-94.

DR PDB; 2OLB; 29-JAN-96.

DR PDB; 1OLC; 29-JAN-96.

DR PDB; 1JET; 15-MAY-97.

DR PDB; 1JEU; 15-MAY-97.

DR PDB; 1JEM; 15-MAY-97.

DR PDB; 1RKM; 29-JUL-97.

DR PDB; 2RKM; 29-JUL-97.

DR PDB; 1B05; 25-NOV-98.

DR PDB; 1B08; 18-NOV-98.

DR PDB; 1B1H; 18-NOV-98.

DR PDB; 1B2H; 18-NOV-98.

DR PDB; 1B3E; 16-DEC-98.

DR PDB; 1B3G; 16-DEC-98.

DR PDB; 1B3H; 18-NOV-98.

DR PDB; 1B3L; 16-DEC-98.

DR PDB; 1B4I; 18-NOV-98.

DR PDB; 1B4Z; 13-JAN-99.

DR PDB; 1B5H; 18-NOV-98.

DR PDB; 1B5I; 13-JAN-99.

DR PDB; 1B5J; 13-JAN-99.

DR PDB; 1B6H; 18-NOV-98.

DR PDB; 1B7H; 25-NOV-98.

DR PDB; 1B9J; 22-FEB-99.

DR PDB; 1B32; 23-DEC-98.

DR PDB; 1B40; 13-JAN-99.

DR PDB; 1B46; 13-JAN-99.

DR PDB; 1B51; 20-JAN-99.

DR PDB; 1B52; 27-JAN-99.

DR PDB; 1B58; 20-JAN-99.

DR PDB; 1QKA; 19-SEP-99.

DR PDB; 1OKB; 09-SEP-99.

DR StyGene; SGI0267; oppa.

DR InterPro; IPR000914; SHP_bac_5.

DR Pfam; PF00496; SHP_bac_5; 1.

DR PROSITE; PS01040; SHP_BACTERIAL_5; 1.

KW Peptide transport; Transport; Periplasmic; Signal; 3D-structure;

KW Complete proteome.

FT SIGNAL 1 26

FT CHAIN 27 543

FT DISULFID 297 443

FT CONFLICT 22 24

FT SEQUENCE 543 AA; 61292 MW; EF344ETC7991CA47 CRC64;

FT PERIPLASMIC OLIGOPEPTIDE-BINDING PROTEIN.

Query Match 15.0%; Score 410; DB 1; Length 543;

Best Local Similarity 26.1%; Pred No. 1.9e-21;

Matches 136; Conservative 97; Mismatches 245; Indels 44; Gaps 15;

QY 4 ISVGICITILLSLSVLQCKESSHSSTSRGELAINTRDEPRSLDPROVRLISEISLVKH 63

Db 10 IAAIG-LTALTAASATAADVPAGVOLADKQTLVNNNGSEVQSLDPHKIEGVPESNVRD 68

QY 64 IYEGVQNNLSGNIPALAEYSSDGLTYTFKLK-SAFWSNGDPLTAEDFIESKQV 122
 Db 69 LFEGLL-ISOVEGHSPGVAEKWE-NKDFKVTWFLHRENAKSDGTPVTAHDFVYSWQRL 126
 QY 123 ATQEVSGIYAFAL-NPIKNVRKIQEGLHSDHFGVHSPNESTLVLTLESPTSHFLKLLA 180
 Db 127 ADPNTASPYASYLOYGHIANIDDIAGKGPATDGLGVAALDDHTFEVTLSEPVYFKLLV 186
 QY 181 LPVFPPVHKS-----QRTLSQSLPIASGAFYKNIKQKWLKSKNHHYNSQVETK 234
 Db 187 HPSVSPVKSAVEKFGDKWTQANI-VINGAYKLKNVWNERIVLERNPQYWDNAKTVIN 245
 QY 235 YITHFTPDANTAAKLFNOGKLNWGPWGERIPQETLSLQSK--GHLHSDVAGTWSL 292
 Db 246 QVYTLPISSSEVTVNRYRSGEIDMT-----YNNMPIELFQKLKEIPNEVRVDPVLCYTY 301
 QY 293 TFNINKPLNMKLRALASALDKEALYSTIFLGRAKTADHLLPTNIHSY----- 342
 Db 302 EINNOKAPFNDVRVTRALKLALDRDIIVNKV---KNQGD--LPA--YSYTPPVTDGAKL 353
 QY 343 --PEHOKQMAORQAVAKKLFKEALEELQITAKDLEHLNLIFFVSSASSLLVQLIREQW 400
 Db 354 VEPEWFKWSQKNEEAKKLAEEA---GFTADKPLTFLLINTSDLHKKLAIAVASIW 408
 QY 401 KESIGRAIPVIGKEFALLQADSSNFSLATGCGWFAEDFADPMFLTIFAYPSGVPPYAIN 460
 Db 409 KKNUGVNVNLENGEMKFTFLDTRHOCTFDVARAGWCADYNEPT'SFLNTMLSDSSNNTAHYK 468
 QY 461 HKDFLEILQIQEODHOKRSELVSQASLYLTFTHIIEPIYH 502
 Db 469 SPAPDKLIADTLKVADDTQRSLEIAKAEQQLDKDSAIIPVY 510
 RESULT 6
 ID YGIS_ECOLI
 AC Q46863; STANDARD; PRT; 535 AA.
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative binding protein ygis precursor.
 GN YGIS OR B3020.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE=99420866; PubMed=10493123;
 RA Pountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
 RT "Enrichment of low abundance proteins of Escherichia coli by
 RT hydroxyapatite chromatography.";
 RL Electrophoresis 20:2181-2195(1999).
 CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM.
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 CC PROTEIN FAMILY 5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U28377; AAA69188.1; -;
 DR EMBL; AE000384; AAC76056.1; -;
 DR HSSP; P06202; 1B52.
 DR EcoGene; EGI3021; ygis.
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 1.
 DR PROSITE; PS01040; SBP_BACTERIAL_5; FALSE_NEG.
 KW Transport; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 535 PUTATIVE BINDING PROTEIN YGIS.
 SQ SEQUENCE 535 AA; 60694 MW; B0F56723071A54EF CRC64;
 Query Match 14.5%; Score 397.5; DB 1; Length 535;
 Best Local Similarity 25.8%; Pred. No. 1.4e-20;
 Matches 126; Conservative 82; Mismatches 233; Indels 47; Gaps 12;
 QY 39 NIRDEPRSLDPRQVRLSEISLVKHYEGLVQENNLGNIEPALAEYSSDGLTYTFK 98
 Db 38 NNHSDPGGLDQPKVEENTAAQIVLDLFEGLVMDG-EGQVQPAQAEWELDGGKRYIFH 96
 QY 99 LKSAF-WNSGDPLTAEDFTESKQVATQEVSGIYA--FALNPIKNVRKIQEGLHSDHFG 155
 Db 97 LRSLGQNSDGOPLTAEDFVLGQRAVDKPTASPFAGYLAQAHINNAAIIVACKADVTSLG 156
 QY 156 VHSNPESLVVLTLESPTSHFLKILALPVFFPV--HKSQRTLSQSLP---IASGAFYPKN 210
 Db 157 VKATDDRTLEVTLEQVPWFVTTMLAWPTLFPVPHVIATKHGDSWSPENWYNGAFVLDQ 216
 QY 211 IQQKWIKLSKNHHYNSQVETKTIITHFIPDANTAALKFNQK--LNWQGPMPGERIP 268
 Db 217 WYNEKITARKNPKYRDAQHTVLQOQVEYALDINSVTGYNRYRAGEVDLTW-----VP 268
 QY 269 QETLSNLQSK--GHLHSDVAGTSMLTENINKFPLNMKLRALASADKALYSTIFLG 326
 Db 269 AQQIPAIEKSLPGLRLIIPRLNSEYINENLEKPPFNDVRVRRALYLTVDRO-LIAQKVLG 327
 QY 327 RAKTADHLLPTNIHSYP----EHOKQMAORQAVAKKLFKEA-----LEELQITAKD 374
 Db 328 LRTATLTTPPEYKGFSAITFDLOKPSERVAMAKALLKQAGYDASHPLRFELFYKNYD 387
 QY 375 LEHLNLIFFPVSSASSLLVQLIREQWESLGFATPIVIGKEFALLQADLSSGNFSLATGGW 434
 Db 388 LHEKTAI-----ALSEWKKWLGAQVTLRTMEWKTYLDARRAGDFMLSRQSW 434
 QY 435 FADFADPMFLTIFAYPSGVPPYAINHKDFLEILQIQEODHOKRSELVSQASLYLTF 494
 Db 435 DATYNDASSFLNTLKSDEENGVHKNQAYDALLNQATQITDATKRNALYQQAEEVINQ 494
 QY 495 HIIEPIYH 502
 Db 495 APLIPIY 502
 RESULT 7
 ID OPPIA_HAEIN STANDARD; PRT; 541 AA.
 AC P71370;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Periplasmic oligopeptide-binding protein precursor.
 GN OPPIA OR H1124.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE OLIGOPEPTIDE
CC PERMEASE, A BINDING PROTEIN-DEPENDENT TRANSPORT SYSTEM, IT BINDS
CC PEPTIDES UP TO FIVE AMINO ACIDS LONG WITH HIGH AFFINITY (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32792; AAC22778.1; -.
DR HSSP: P06202; IBS2.
DR TIGR: H1124; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 1.
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
KW Peptide transport; Transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 541 PERIPLASMIC OLIGOPEPTIDE-BINDING PROTEIN.
SQ SEQUENCE 541 AA: 60906 MW; 1B949C9617514A99 CRC64;
-----
Query Match
Best Local Similarity 13.5%; Score 370; DB 1; Length 541;
Matches 132; Conservative 97; Mismatches 235; Indels 64; Gaps 18;
QY 10 ITILLSL-----VVLQCK--ESSHSSTSRGELAINIREPSLDRQVRLISELSLVKH 63
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 IALALSYSAQAVIVPEGTQLDEKQ-----IVINNGAEPSQSPDPHKTGVPESNVAQ 62
QY 64 IYGLVOENLNSGTEPALAEDYSLSDGLTYTFKL-KSAFWSNGDPLTAEDFIESHKQV 122
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 LLEGVLTSDS-EGKLQPGAESWENTPDKTWTFHLRKDAKNSGDPVTAHDFVFAWRRL 121
QY 123 ATQEVSGIVAFALN--PKNVKRTQEGHLSIDHFGVHSPNESTLVLTLESPTSFL---- 176
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 VDPATAPYASLYSLQVENAQDILDGKKPAELGVEXAKDDYTFVHTTNVPTVTSXYT 181
QY 177 --KLALPVEFFPVHKSQRTLSKSLPITASGAFYPKNIKQKOWIKLSKNPHYNSQVETK 234
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 HQSLPLPKVY-VEKLGDAWKYKENVYNGAYKLANHIENKIEFERNPLVNDKETVIN 240
QY 235 TTIHTIPDANTAALKFNQGLNMQGPWGERIPQETLSNLSQK--GHLSFDVAGTSWL 292
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 SATFLAENPSTDVARYRAGLDWTS--YG--LPPEQFAKLQKEPLGVEVYVTRTLQTSY 296
QY 293 TFINKFPNNKLEALASALDEKALVSTIFLGRAKTADHLL-PTNIH-----SYPEHQ 346
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
297 ELANKAPFDNINIRKALNSLDNRNVTIDKV-LGQQTPTVFTYIEBGLIQQPAYTS 355
QY 347 KOEMAQORAYAKKLFKEA-----LEEIQITAKLEHLNLIIPVSS-----SASSLL-VQ 394
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
356 KEPMARNEALKLEEGAGSKANPLKFSILYNTNENHKKVAIAAASMMWKANTKGLIDVK 415

```

```

QY 395 LIREQWKESLGFAIPVIGKEFALLQADLSSGNFSLATGWFADFPADMAFLITFYAPSGV 454
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
416 LENQWKTYID-----SRRAGRYDVARAGNADYNQATTFNGYFELSSEN 460
QY 455 PPYAIHKHDFLEILQNIQEOHQKRSLSVQSASLYLSETHIEPIYH 502
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
461 NTAKYANPEYDKAMBSYAATDAEGRAKAYAKAEILGKDYGVIPFN 508
-----
RESULT 8
Y213_HAEIN STANDARD; PRT; 514 AA.
ID Y213_HAEIN
AC P44572;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative binding protein HI0213 precursor.
GN HI0213.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32706; AAC21881.1; -.
DR HSSP: P06202; IBS2.
DR TIGR: HI0213; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
KW Hypothetical protein; Transport; Membrane; Signal; Lipoprotein;
KW Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 514 PUTATIVE BINDING PROTEIN HI0213.
FT LIPID 24 24 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 514 AA: 58876 MW; 789188C4328BDEBC CRC64;
-----
Query Match
Best Local Similarity 12.9%; Score 354.5; DB 1; Length 514;
Matches 128; Conservative 105; Mismatches 213; Indels 83; Gaps 20;

```



```

Db 460 VLGSNDVVI--DIQQLTSDFSGYFAQTAQKDYLDYHGGWGPYQDPSTYLDIENFN 517
QY 452 S-----GVPPYAINHKD-----FLEILONTEQDQKRSKSELVSOASLYLETFHII 497
Db 518 SGGFLQNLGLEPCEANDKAKAVGLDVYTMLEANEKEQDPKARYEKYADIQAWLIDSSLV 577
QY 498 EPIYHDAFOFAMKKL---SMLGVSPY-GVVDFERYAK 530
Db 578 LPSVSRGGTPSLRRVTPFAAAYGLTGKGVESYKYLK 614

RESULT 10
NIKA_ECOLI
ID NIKA_ECOLI STANDARD; PRT; 524 AA.
AC P33590;
DT 01-FEB-1994 (Rel. 28, Created)
DE 16-OCT-2001 (Rel. 28, Last sequence update)
DE Nickel-binding periplasmic protein precursor.
GN NIKA OR B3476.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95020649; PubMed=7934931;
RA Navarro C., Wu L.-F., Mandrand-Berthelot M.-A.;
RT "The nik operon of Escherichia coli encodes a periplasmic binding-
RT protein-dependent transport system for nickel.";
RL Mol. Microbiol. 9:1181-1191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- REPRESENTS INVOLVED IN A NICKEL TRANSPORT SYSTEM, PROBABLY
CC -!- SUBCELLULAR LOCATION: Periplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73143; CA851659.1; .
DR EMBL; U00039; AAB18451.1; .
DR EMBL; AE000423; AAC76501.1; .
DR PIR; S39594; S39594.
DR EcoGene; EG12075; nika.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
DR Transport; Nickel; Signal; Periplasmic; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 524 NICKEL-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 524 AA; 58719 MW; CB2E3C1DCBE42396 CRC64;

Query Match 11.0%; Score 301; DB 1; Length 524;
Best Local Similarity 24.2%; Pred. No. 8, 4e-14;
Matches 119; Conservative 93; Mismatches 209; Indels 70; Gaps 18;

QY 64 IVEGLVQNNLSGNIPEALAEYSLSDDGLTYTFKLK-SAFWSNGDPLTAEDFIESWKQV 122

```

```

Db 55 VVEPLV-KYQADGVSFWLAKSWTHSEDKTWTFLRDVVKFSGNPFDAEAAENFRAV 113
QY 123 ATQEVSGIYAFALNPINKVRKIQEGHLSIDHFGVHSPNESTLVVLTLESPTSHFLKLALP 182
Db 114 LDNRORHAWLELANQIVDKAL-----SKTELQITLKSAYYFFLOELALP 158
QY 183 VFF-----PVHKSORTLOSXSUPIASGAFYPKNIKOKWIKLSKNPHYTNOSQVETKIT 237
Db 159 RPFRIAPSOFKHETMNGIKAPIGTGPWILQESKLNQYDVFVRNENYWGKPA-IKKIT 217
QY 238 IHFIPDANTAAKLFNOGKLNWOGPPWGER--IPQETLSNLSQSKGHLHS--FDVAGTSWLT 293
Db 218 ENVIPDPTTRAVAFETGDIIDL---YGNELLLPLOTFAFSONPAYHTOLSQPIETVMLA 274
QY 294 FNINKFPLNNKRLREALASALDKALVSTIFLGRAKTADHLLPTNI-----HSYPEH 345
Db 275 LNTAKAPTNELAVREALYVAVNKSLIDNALYGTQQVADYLFAPSVPYANGLKPSOYDP 334
QY 346 QKQEMARQAYAKKLFKEALELQITAKDLEHLN-----LIFFVSSSASSLLVQLIRE 398
Db 335 QK-----AKALLEKAGWTLF-AGKDIREKNGOPLRIELSTFGTDALSKSMAEIIQA 384
QY 399 QWKESLGAIPIVGKEFALLQADLSSGNFSLA-TGGWFADFPDPAFLTIFAYPS----- 452
Db 385 DMQ-IGADVSLIGEEESIYARQDRGFNFHRTWGAPY-DPHAFSSMRVPSHADFO 442
QY 453 --GVPPYAINHKDFLEILQNEQDQKRSSELVSOASLYLETFHIIETIYHDAFOFAM 509
Db 443 AQOGLADKPLIDKEIGEVLAETHDTQQAALYRLILTR--LHDEAVYL--PLSYISMVMVVS 498
QY 510 NKKLSNLGVSP 520
Db 499 KPELGNIPYAP 509

RESULT 11
HBPA_HAEIN
ID HBPA_HAEIN STANDARD; PRT; 547 AA.
AC P33950;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme-binding protein A precursor (Hemin-binding lipoprotein).
GN HBPA OR DHPA OR HI0853.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DL42 / SEROTYPE B;
RX MEDLINE=92267636; PubMed=1339409;
RA Hanson M.S., Slaughter C., Hansen E.J.;
RT "The hbpA gene of Haemophilus influenzae type b encodes a
RT heme-binding lipoprotein conserved among heme-dependent Haemophilus
RT species.";
RL Infect. Immun. 60:2257-2266(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";

```

RL Science 269:496-512(1995).

RP [3] POSSIBLE FUNCTION, AND SUBCELLULAR LOCATION.

RC STRAIN-DL42 / SEROTYPE B;

RX MEDLINE=91251755; PubMed=2041470;

RA Hanson M.S., Hansen E.J.;

RT "Molecular cloning, partial purification, and characterization of a

RL haemin-binding lipoprotein from *Haemophilus influenzae* type b.";

RT Mol. Microbiol. 5:267-278(1991).

CC -!- FUNCTION: IMPORTANT ROLE IN HEME ACQUISITION OR METABOLISM.

CC -!- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid anchor.

CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M88134; AAA73214.1; ALT_SEQ.

DR EMBL; M84028; AAA24962.1; .

DR EMBL; U32767; AAC22512.1; ALT_INIT.

DR PIR; A43832; A43832.

DR HSP; P23847; LDPE.

DR TIGR; H10853; .

DR InterPro; IPR000914; SBP_bac_5.

DR Pfam; PF00496; SBP_bac_5; 1.

DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.

FW Inner membrane; Signal; Lipoprotein; Complete proteome.

KW SIGNAL; 18 PROBABLE.

FT CHAIN 19 547 HEME-BINDING PROTEIN A.

FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).

FT VARIANT 48 49 KA -> NS (IN STRAIN DL42).

FT VARIANT 181 181 T -> N (IN STRAIN DL42).

FT VARIANT 240 240 H -> N (IN STRAIN DL42).

FT VARIANT 343 343 T -> I (IN STRAIN DL42).

FT VARIANT 375 375 A -> V (IN STRAIN DL42).

SQ SEQUENCE 547 AA; 60660 MW; 7755E2164A40CB31 CRC64;

Query Match 10.6%; Score 291.5; DB 1; Length 547;

Best Local Similarity 23.1%; Pred. No. 4.2e-13;

Matches 128; Conservative 95; Mismatches 259; Indels 73; Gaps 18;

QY 13 LLSLSVLQCKSSHS-----STSRGELAINIRDEPRSLDPRO-VLLSEISLV 61

DB 9 LAATLVLAACDQSSANKSTAQTEAKSSNNVTVCYCTAKPLGFSPLITRGTSYNASS 68

QY 62 KHIEGLVQVNNLSGNIIEPALAEDYSLSSDGLATYFKLSA-----FWSNGDPLTAE 113

DB 69 QQVNRVLVEFKGSTDIEPALAESWEISDGLSVTFHLRGVGFHTYFTPTD-FNAD 127

QY 114 DFIESKQVATQEVSGIYAFALNPKIKVNRKIQEGLSIDHF-----GVHSPNESTLVTL 168

DB 128 DWVFSFQRLDPN-----HPYHNVSGTYPYFKAMKFPPELLKSVKVDNTITIL 178

QY 159 ESPTSFLKLLALPVFPVHKSQ-----RTLQSKSLPIASGAFYKNIKQKWI 217

DB 179 NKTDATPLASLGMDFISYSAEYADSLKAGKPELDSR--PVGTGPFVFDVYKTDQAI 235

QY 218 KLSKNPHYNQSOVKTKITHTIPDANTAALFNQGLNW-QGPPWGERIPQET---LS 273

DB 236 QYVAHENYW-KGRTPDLRLVISIVPDATTRYAKLQAGTCDDLLFPNADIAKMKRTPKVQ 294

QY 274 NLQSKGLHLSFDVAGTSWLTNTINKFPLNNKRLREALSALDKALYSTIFLGRKATADH 333

DB 295 LLEQKG-----LNVAITAEKAFPDNVKVRQALNVAVDKKAITEVYQAGTSASN 347

QY 334 LLPTNIHSPHQEQMAQROAVAKKLEALEELQITAKDLHLNLFVY-----SSASS 390

DB 348 PLPPTIWSYND-EIQDYPYDPEKAKQLLAE-----GYPNGFTDFQIWPVIRASNPNK 401

QY 391 LLVQLIRQWESLGFPIPIVGKFEALLQADLSSNFSLATGGMFADFADPMAFLTIPAY 450

DB 402 RMAELITMDWAK-IGVKTNPVTYEWADYRKRAKEGELTAGIFGWSGSDGDPDNFLSPILG 460

QY 451 PSGV---PPYAINHKDELEILQINIEQDQHKRSELVSQASLYLETFHIEPIYHDAFQF 507

DB 461 SSNIGNSNWARFNSEFDALLNEAIGLTNKEAKLYKQAOVIVHQAQWIPVAHSVGF 520

QY 508 ANKKLSNLGVSPG 522

DB 521 PLSPRVKGYVQSPG 535

RESULT 12

YDDS_ECOLI

ID YDDS_ECOLI STANDARD; PRT: 516 AA.

AC P76128; P77769; P76874;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Putative ABC transporter periplasmic binding protein ydds precursor.

GN YDDS OR B1487.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RT "The complete genome sequence of *Escherichia coli* K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12.

RX MEDLINE=97251357; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Tagami H., Takeda J., Takeuchi K., Takeuchi Y., Wada C., Yamanoto Y., Horiuchi T.;

RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM YDDOPQRS.

CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).

CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AE000245; AAC74560.1; .

DR EMBL; D90789; BAA15142.1; ALT_INIT.

DR EMBL; D90790; BAA15152.1; ALT_INIT.

DR EMBL; D90791; BAA15158.1; ALT_INIT.

DR EcoGene; E813790; ydds.

```

DR InterPro; IPR000914; SBP_dac_5.
DR Pfam; PF00496; SBP_dac_5; 1.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW Hypothetical protein; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 516 PUTATIVE ABC TRANSPORTER PERIPLASMIC
FT SEQUENCE 516 AA; 57641 MW; 86E1C32CC3E06FB9 CRC64;
SQ

Query Match          10.5%; Score 287.5; DB 1; Length 516;
Best Local Similarity 22.1%; Pred. No. 7.3e-13;
Matches 123; Conserved 109; Mismatches 208; Indels 117; Gaps 20;

QY 16 LSVVLOGKESHSSTSGELAINIRDEPRSLDPROVRLLESEISLVKHIYEGLVQ----E 71
DQ 12 LALVLAETNPAAHAAVPMKMLVIGKAAADPOTLDPAVTIDNDWTVPVSORLVQKTDG 71
QY 72 NNLGNIPEALAEYSSDGLATYFKLK-SAFWSNGDPLTAEDFIESKQVATQEVSGI 130
DQ 72 DKGSTDVEGLASSKASDDQKWTFTLKDNAKFAAGTPTVAEAVKLSFERLL----- 124
QY 131 YAFALNPIKNVRKIQEG-----HLSIDHFGVHSPNESTLWLTLESPTSHFLLKLA--- 180
DQ 125 -----KIQGPAAEFKPKLKID-----APDEHTVKFTLSQTPAFPLYTLANDG 167
QY 181 ----LPVFFPVHK-----SORTLQSKSLPIASGAFYKPKIKQKQWIKLKNPHYNYQ 228
DQ 168 ASIINPAVLKEHAADDARGFLAQT-----AGSGPFMLKSKWQKQQLVLPNPHYPG- 219
QY 229 SOVETKTIITHIFDANTAALFNQCKLNWQGPWGERIPQETLSNL--QSKGHLSFDV 286
DQ 220 NKPNEKRVSVKIKGESASRLQSLRGDID-----IADALPVDQLNALKQENKVNVAEYPS 274
QY 287 AGTSLMIFNTNKPFLNMKLALASALDKALYSTTFLGRATADHLLPTNIHSYBEHQ 346
DQ 275 LRVTYLVLNNSKAPLQADLRASISWSDYOGVNGILSGKQMGPIEGMGYD--- 331
QY 347 KOEAAQCAQAKKLFKEALEEIQITAKDEHLNLF-----PVSSASLLVQLIRE 398
DQ 332 ----ATAMQVNHDETKAKAEWKVTSKPTS-LTFLYSDNDPWEPIALATOSSL----- 380
QY 399 QWESLGFALPIVGEFALLQADLSSGNFSLATGCGWADPADPMAFLTIF--AYPSGVPP 456
DQ 381 ----NKLIIIVKLEKLANATMDRVKGGDYDIAIGNNSPDPADPYEMMYWESDKGLP- 436
QY 457 YAINHKDFLE-----ILQN-----IEQEDHOKRSELVSQASLYLETHIETPIYHD 503
DQ 437 ---GNRSFVENSEVDKLLRNALATTDQTRDYQQAQKIVIDDAAYV-----LFQK 486
QY 504 AFQFAMNKLNLGVSP 520
DQ 487 NYQLANKEKGVGFNP 503

RESULT 13
DPPA_ECOLI
ID DPPA_ECOLI
AC P23847; STANDARD; PRT; 535 AA.
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Periplasmic dipeptide transport protein precursor (Dipeptide-binding
DE protein) (DBP).
GN DPPA OR B3544..
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```

```

RX MEDLINE=911100289; PubMed=1702779;
RA Olson E.R., Danyak D.S., Jurss L.M., Poorman R.A.;
RT "Identification and characterization of dppA, an Escherichia coli
RT gene encoding a periplasmic dipeptide transport protein.";
RL J. Bacteriol. 173:234-244(1991).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-50.
RC STRAIN=K12;
RX MEDLINE=92065799; PubMed=1956284;
RA Abouhamad W.N., Manson M., Gibson M.M., Higgins C.F.;
RT "Peptide transport and chemotaxis in Escherichia coli and Salmonella
RT typhimurium: characterization of the dipeptide permease (Dpp) and the
RT dipeptide-binding protein.";
RL Mol. Microbiol. 5:1035-1047(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MM500;
RX MEDLINE=95231288; PubMed=7536291;
RA Abouhamad W.N., Manson M.D.;
RT "The dipeptide permease of Escherichia coli closely resembles other
RT bacterial transport systems and shows growth-phase-dependent
RT expression.";
RL Mol. Microbiol. 14:1077-1092(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [5]
RP SEQUENCE OF 29-40.
RC STRAIN=K12 / BMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robinson K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [6]
RP DISULFIDE BONDS.
RA Duntzen P.;
RL Submitted (JUN-1995) to the SWISS-PROT data bank.
RN [7]
RX X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
RA Duntzen P., Nowbray S.L.;
RT "Crystal structure of the dipeptide binding protein from Escherichia
RT coli involved in active transport and chemotaxis.";
RL Protein Sci. 4:2327-2334(1995).
RN [8]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RA MEDLINE=96118375; PubMed=8527431;
RA Nickitenko A.V., Trahanov S., Quilcho F.A.;
RT "2-A resolution structure of dppA, a periplasmic dipeptide
RT transport/chemosensory receptor.";
RL Biochemistry 34:16585-16595(1995).
CC -!- FUNCTION: DIPEPTIDE-BINDING PROTEIN OF AN OSMOTIC-SHOCKABLE
CC TRANSPORT SYSTEM. DPPA IS ALSO REQUIRED FOR PEPTIDE CHEMOTAXIS.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35045; AAA23707.1; -.
CC EMBL; X58051; CAA41090.1; -.

```

DR EMBL; L08399; AAA23702.1; -
 DR EMBL; U00039; AAB18522.1; -
 DR EMBL; AE000431; AAC76569.1; -
 DR PIR; S15292; S15292.
 DR PIR; A39194; A39194.
 DR PDB; 1DPE; 17-AUG-96.
 DR PDB; 1DPP; 17-AUG-96.
 DR SWISS-2DPAGE; G059.9; 6TH EDITION.
 DR EMBL; E010248; dppA.
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 1.
 KW PROSITE; PS01040; SBP_BACTERIAL_5; 1.
 KW Peptide transport; transport; periplasmic; signal; Chemotaxis;
 KW 3D-structure; Complete proteome.
 FT SIGNAL 1 28
 FT CHAIN 29 535 PERIPLASMIC DIPEPTIDE TRANSPORT PROTEIN.
 FT DISULFID 34 262
 FT DISULFID 450 463
 SQ SEQUENCE 535 AA; 60293 MW; 3F7C4756E2A8C2F0 CRC64;

Query Match 10.4%; Score 283.5; DB 1; Length 535;
 Best local similarity 24.6%; Pred. No. 1.5e-12;
 Matches 122; Conservative 85; Mismatches 223; Indels 65; Gaps 20;

QY 64 IYELGVQENLSCNIEPALAEYLSLSDGLTYFKL-KSAFWSNGD-----PLTAEDFI 116
 Db 60 LYNLVEFKIGTEVTEVPLGAKEVSESDGKTYTFHLRGVKNHDKFEKPTRELNADVV 119
 QY 117 ESMQVATQVSGYIYAFALPIKVRKIQEGHLSIDHP-GVHSP-----NESTLVV 166
 Db 120 FSDRQKN-----AQNP---YHKVSGG---SYEYFEGMGLPELISEVKVVDNTQVF 165
 QY 167 TLESTSTHFLKLLALPVFFVHKRSORTLQSKS-----LPTASGAFYPKNIKQKQIK 218
 Db 166 VLTREAPFLADLANDFASILSEYADAMKAGTPEKLDLNPITGTFPQLQYQKDSRIR 225
 QY 219 LSKNPHYV-NGSOVETWITIHIPDANTAKLFNQK---LWQGPFPNGERIPQETLSN 274
 Db 226 YKAFQGVNGTKPQDITLVFSI---TPDASVRYAKLQKNEQVMPYNPADIARMKQDKIN 283
 QY 275 LOSGHLHSFDVAGTSMIFNFKPLNNMKLREALASALDKALYSTIFLGRKATADHL 334
 Db 284 LMEMPGLN-----VGYSYNNQKPLDDVYKQALYVNNKDAIKAYVQAGVSAKNL 337
 QY 335 LPTNHSYPHQEQAQMAQYAKLKFKEALELQITAKOLEHLNIFPVSSASSLLVQ 394
 Db 338 IPTPMWGYND-DVQDYTYDPEKAKALLKEAGLEKGEFSI-DLWAMPVQRYNPNNRR-MAE 394
 QY 395 LIRQWRESGLGFAIPVIGKEFALLQADLSSGNFSLATGWFADFPD-MAFLTTF---AY 450
 Db 395 MIQADWAK-VGVQAKVTVWEGEYKRAKGEHQTVMMGWTGNDGDPNFFATLFSKAAS 453
 QY 451 PSGVPPVAINKHDFLEILQNIQEODHQRSELSQASLYLETFHILP---IYHDAQF 507
 Db 454 EQGSNYSKWCYKPFEDLIQPARATDDHKNRVLYKQAVVM---HDQAPALIIAHSTVFE 510
 QY 508 AMNKLNLGVSPGT 522
 Db 511 PVREKVGWYDPLG 525

RESULT 14
 ID AMIA_STRPN
 AC AMIA_STRPN STANDARD; PRT; 659 AA.
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Oligopeptide-binding protein amIA precursor.
 GN AMIA OR SPI891.
 OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 RX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R800;
 RX MEDLINE=90279506; PubMed=2352474;
 RA Allouing G., Trombe M.C., Claverys J.-P.;
 RT "The ami locus of the Gram-positive bacterium Streptococcus pneumoniae
 is similar to binding protein-dependent transport operons of gram-
 RT negative bacteria";
 RL Mol. Microbiol. 4:633-644(1990).
 RN [2]
 RP REVISIONS.
 RA Claverys J.-P.;
 RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae";
 RL Science 293:498-506(2001).
 RN [4]
 RP SEQUENCE OF 1-28 FROM N.A.
 RC STRAIN=R800;
 RX MEDLINE=90060772; PubMed=2684766;
 RA Martin B., Allouing G., Boucraut C., Claverys J.-P.;
 RT "The difficulty of cloning Streptococcus pneumoniae mal and ami loci
 in Escherichia coli: toxicity of malX and amiA gene products.";
 RL Gene 80:227-238(1989).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=89091147; PubMed=3208757;
 RA Gilson E., Allouing G., Schmidt T., Claverys J.-P., Dudler R.,
 RA Hofnung M.;
 RT "Evidence for high affinity binding-protein dependent transport
 systems in Gram-positive bacteria and in Mycoplasma";
 RL EMBO J. 7:3971-3974(1988).
 CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
 ANCHOR.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 PROTEIN FAMILY 5.
 CC -!- CAUTION: THE REVERSED SEQUENCE OF AMIA NOW INCLUDES, IN THE
 C-TERMINAL SECTION, THE SEQUENCE OF AN ORF WHICH WAS PREVIOUSLY
 KNOWN AS AMTB.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; X17337; CAA35213.1; -
 DR EMBL; AE007479; AAK75962.1; -
 DR PIR; S11148; S11148.
 DR PIR; S11149; S11149.
 DR TIGR; SPI891; -
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

DR PROSITE; PS01040; SBP_BACTERIAL5; 1.
 KW Peptide transport; Transport; Signal; Membrane; Lipoprotein;
 FT Complete proteome.
 FT SIGNAL 1 22 PROBABLE.
 FT CHAIN 23 659 OLIGOPEPTIDE-BINDING PROTEIN AMIA.
 FT LIPID 23 23 N-ACYL DIGLYCERIDE (PROBABLE).
 FT CONFLICT 54 54 S -> R (IN REF. 1).
 FT CONFLICT 421 421 A -> R (IN REF. 1).
 FT CONFLICT 557 557 R -> K (IN REF. 1).
 FT CONFLICT 562 562 L -> F (IN REF. 1).
 FT CONFLICT 622 622 S -> L (IN REF. 1).
 FT CONFLICT 626 626 V -> A (IN REF. 1).
 SQ SEQUENCE 659 AA; 72465 MW; C2ED06A1DB8A7B81 CRC64;

Query Match 10.18; Score 275.5; DB 1; Length 659;
 Best Local Similarity 21.68; Pred. No. 7.3e-12;
 Matches 140; Conservative 107; Mismatches 255; Indels 147; Gaps 26;

QY 1 MRKISVIGICITILLSVLOGCKESHSTSRGELAIN--IRDEPSRLDPQVRLLEI 58
 Db 1 MKNRVFATAGLVLLAAGVLAAC--SSKSSDSAPKAYGYVTADPETLDYLLSSKNSTT 59
 QY 59 SLVKHYEGLVGNLSNIEPALAEYSLSSDGLTYTFKLKSA---FWNGD---PLTA 112
 Db 60 VVTSNGIDGLTNDY--GNLAPVAEDWEYSGDGLTYTKIRKGVKWFSDGEYAEVTA 118
 QY 113 EDFIESKQAVQEVSGYIFALNPINKVKTQEG--HLSIDHFGVHSPNESTLWVLESP 171
 Db 119 KDFVGLKHAADKSEAY--LAENSVKGLADYLSGTSTDFSTGVGVKAVDDYTLQYTLNQP 177
 QY 172 TSFHLKLLALPFFPVHKRSORTLOSKSLP-----IASGAFPKNIKOKWIKLSKP 223
 Db 178 EPFWNSKLTYSIFWPLNEEFETSKGSDFAKPTDPTSLYNGPFFLLKGLPAKSSVEFVKN 237
 QY 224 HYVNSQVETTYI-----HFIPDNTAAKLFNOCKLNWQGPWGERIPQE 270
 Db 238 QVWDENVHLDITINLAVYDGSQESLERNFSTGAYSVARLYPTSS--NY-----SKVAEE 290
 QY 271 TLSNLSQKGLHSPDVAGTSWLTFFNINKPFLN-----NMKLEALAS 312
 Db 291 YKDN1-----YITGSGSGLAGLVNDROSYNVTSKTTDSEKVAATKALKINDFRQALNF 345
 QY 313 ALDKEA-----LVSTFVL-----GRAKTADHLLPTNIHSYPEHOK----- 347
 Db 346 ALDRSAVSAQINGKGAALAVRLNLFVKPDPVSAGEKTFGLVAQAQIPAYGVDEKGVNLAD 405
 QY 348 -QEMAQROAVAKLKEALEELQITAKDLEHLNLFIPVSSASSLL--VOLIREQWKESL 404
 Db 406 QGDGLFNADKAKAEPAKAKALEADGVQFP--IHLDPVPDQASKNYISRTQSFQSVETVL 464
 QY 405 GFAPIV-----GKEF---ALLQADLSGNFSLATG--GWFAFDADPMAFLTI----- 447
 Db 465 GVENVVDIQQMTSDEFNLITYYANASSEDWDVSGVSGWGPDYQDPSTYLDLTKTSS 524
 QY 448 -----FAYPSVPPYAINKHDFLEILQNTQEG--QDHQKRSYLSQASLYLEFTHIEP 499
 Db 525 TTKYLLGDFNPSPVQVGLKEDYKLVDEAARETSDLNVRVEKYAAQAQWLTSSIFIP 584
 QY 500 IYHDAFOFAMNKLNLGVSPP-----TGCV-----VDFRYAK 530
 Db 585 -----AMASSGAAPVLSRIVPFTGSAQTGSKGSDVYFKYLK 621

RESULT 15
 ID ALIA_STRPN
 AC P35592; 054782; 054620; 052228; PRT; 660 AA.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Oligopeptide-binding protein alia precursor (Exported protein 1).
 GN ALIA OR EXP1 OR PLPA OR SP0366.

OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_taxid-1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R800;
 RX MEDLINE=94328326; PubMed=8051706;
 RA Alloing G., de Philip P., Claverys J.-P.;
 RT "Three highly homologous membrane-bound lipoproteins participate in
 oligopeptide transport by the Ami system of the Gram-positive
 Streptococcus pneumoniae.";
 RL J. Mol. Biol. 241:44-58(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCITC 11906, SP-496, SP-VA92, AND SP-VA96;
 RX MEDLINE=98125733; PubMed=9466257;
 RA Coffey T.J., Enright M.C., Daniels M., Morona J.K., Morona R.,
 Rykiewicz W., Paton J.C., Spratt B.G.;
 RT "Recombinational exchanges at the capsular polysaccharide
 biosynthetic locus lead to frequent serotype changes among natural
 isolates of Streptococcus pneumoniae.";
 RL Mol. Microbiol. 27:73-83(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae.";
 RL Science 293:498-506(2001).
 RN [4]
 RP SEQUENCE OF 18-660 FROM N.A.
 RC STRAIN=R6X;
 RX MEDLINE=95020610; PubMed=7523829;
 RA Pearce B., Naughton A.M., Masure H.R.;
 RT "Peptide permeases modulate transformation in Streptococcus
 pneumoniae.";
 RL Mol. Microbiol. 12:881-892(1994).
 RN [5]
 RP SEQUENCE OF 347-509 FROM N.A.
 RC STRAIN=R6X;
 RX MEDLINE=95020625; PubMed=7934910;
 RA Pearce B.J., Yin Y.B., Masure H.R.;
 RT "Genetic identification of exported proteins in Streptococcus
 pneumoniae.";
 RL Mol. Microbiol. 9:1037-1050(1993).
 CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (Probable).
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 PROTEIN FAMILY 5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z35135; CA84507.1; -;
 DR EMBL; AE007348; AA874534.1; ALT_INIT.
 DR EMBL; AF030359; AAC38676.1; -;
 DR EMBL; AF030360; AAC38681.1; -;

DR EMBL; AF030361; AAC38686.1; -
DR EMBL; AF030364; AAC38703.1; -
DR EMBL; L20556; AAA26952.1; -
DR TIGR; SP0366; -
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW Peptide transport; Transport; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 22 PROBABLE.
FT CHAIN 23 660 OLIGOPEPTIDE-BINDING PROTEIN ALIA.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (PROBABLE).
FT VARIANT 7 7 F -> L (IN STRAIN R800).
FT VARIANT 27 27 G -> D (IN STRAINS NCTC 11906, SP-VA92,
FT SP-496 AND SP-VA96).
FT VARIANT 30 30 T -> A (IN STRAINS R800 AND R6X).
FT VARIANT 166 166 E -> D (IN STRAINS NCTC 11906, SP-VA92,
FT SP-496 AND SP-VA96).
FT VARIANT 246 246 V -> I (IN STRAINS R800 AND R6X).
FT VARIANT 368 368 L -> I (IN STRAINS NCTC 11906, SP-VA92,
FT SP-496 AND SP-VA96).
FT VARIANT 432 432 Q -> T (IN STRAINS R800, R6X, NCTC 11906,
FT SP-VA92, SP-496 AND SP-VA96).
FT VARIANT 559 559 T -> A (IN STRAINS NCTC 11906, SP-VA92,
FT SP-496 AND SP-VA96).
FT VARIANT 612 612 T -> A (IN STRAIN SP-496).
FT CONFLICT 18 19 TT -> GV (IN REF. 4).
FT CONFLICT 137 137 L -> P (IN REF. 4).
FT CONFLICT 420 420 A -> R (IN REF. 1).
SQ SEQUENCE 660 AA; 73079 MW; 702556f92ec0558 CRC64;

Query Match 9.9%; Score 272.5; DB 1; Length 660;
Best Local Similarity 21.4%; Pred. No. 1.2e-11;
Matches 128; Conservative 105; Mismatches 229; Indels 137; Gaps 24;

QY 12 ILISLSVVGKSGSSSTSRGELAINIRDEPRSLDPQVRLLSEISLVKHYEGLYOE 71
DB 12 VILLAAATLACSGSGSTGKGTFTSYIYETDPDNLNLTAKAANTANITSNVVDGLL-E 70
QY 72 NNLSSNIEPALAEDYLSDDGLTYFKL-KSAFW--SNGD---PLTAEDFIESWKQVATQ 125
DB 71 NDYGNFVPSMAEDWSVKDGLTYTIRKDAKWYTSEGEYAAVRAQDFVTGLKYAADK 130
QY 126 EYSGIYAFALNPINKVKRTOEHL-SIDHFGVHSPNESTLVVLTESPTSHFLKLLALPVF 184
DB 131 KSDALY-LVQESINGLDAYVKEIKDFSGVGIKALDEQTVQYTLNKPESFWSKTTMGVL 189
QY 185 FVHKSQRTLSQK-----SLPIASGAFYPKNIKQKWLKSKNPHYNSQVETK 234
DB 190 APV--NEEFLNSKGDGFAKATDPSSLLYNGPYLLKSIVTKSSVEFAKNPNYWDKNHVD 247
QY 235 TITHTFT--PDANTAALFNQGLKNWGPWPNGERI-----AARLYPTSASFAELEKSKMKNIVYTQDS 271
DB 248 KYKLSFWGDGDTSKPAENEFKDGSLT-----AARLYPTSASFAELEKSKMKNIVYTQDS 301
QY 272 L-----SNLQSKGHLSPDVAGTSHLFTNFKPLNNMKLREALASALDKALYS----- 321
DB 302 IYLVGINIDRQSYKT-----SKTDEQKASTKALLNKDFQAIAPGFDRTAYASQLNGQ 358
QY 322 -----TIFLGRKATADHLPTNIHYSPEHOK-----QEMAQROAYAKKL 360
DB 359 TGASKILRNLFVPTFVQADGKFCDMVKKELTYGDEWKNVDNLSQDGLYNPEKAKAE 418
QY 361 FKEALELOITAKDLEH-LNLIFFVSSASLL--VOLIREQWKEISLG----- 405
DB 419 FAKAKSALQ--AEGVQFPIHLDMPVDQATTQVQRVQSMKQSLKATLGADNVIIDIOQLQ 476
QY 406 -----FAIPVIGKEFALLQADLSSGNFSLATCGWFADPDMAPL----- 445
DB 477 KDEVNITYFAENAGEDW-----DLSD-----NVGWPDPFADPSTVLDIIRPSVGEST 525

QY 446 -TIFAYPSG---VPPYALNHKDFLEILQNT-BQEOHOKRSELYSQASLYLETFHIEP 499
DB 526 KTVLGFDSGEDNVAKKVGLDYDEKLVTEAGDETTDVAKRYDKYAAAAQAWLTDSALIIP 584

Search completed: July 26, 2002, 04:42:32
Job time: 260 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 04:36:22 ; Search time 87.81 seconds

(without alignments)

1048.096 Million cell updates/sec

Title: US-09-824-567-2

Perfect score: 2739

Sequence: 1 MKKISVIGITILLSVVL.....LSNLGVSPTGVWDFRYAKEN 532

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriopl:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2739	100.0	532	16 Q928Y9	Q928Y9 chlamydia p
2	1364	49.8	518	16 Q84201	Q84201 chlamydia t
3	1345	49.1	520	16 Q9PKJ4	Q9PKJ4 chlamydia m
4	1160.5	42.4	527	16 Q928Y8	Q928Y8 chlamydia p
5	755.5	27.6	526	16 Q928Y6	Q928Y6 chlamydia p
6	755.5	27.6	528	16 Q9K244	Q9K244 chlamydia p
7	748	27.3	527	16 Q9PKL9	Q9PKL9 chlamydia m
8	685.5	25.0	529	16 Q84178	Q84178 chlamydia t
9	536.5	19.6	550	16 Q97046	Q97046 clostridium
10	509.5	18.6	528	2 Q9K5F7	Q9K5F7 borrelia ga
11	506	18.5	528	2 Q9K5J9	Q9K5J9 borrelia af
12	505.5	18.5	528	2 Q9K5J9	Q9K5J9 borrelia bu
13	505.5	18.5	528	16 Q54584	Q54584 borrelia bu
14	482.5	17.6	435	16 Q928Y7	Q928Y7 chlamydia p
15	479.5	17.5	532	16 Q9AKR0	Q9AKR0 rhizobium m
16	478	17.5	559	16 Q929H6	Q929H6 listeria in

17	476	17.4	550	2 Q93QH8	Q93QH8 lactococcus
18	474.5	17.3	558	2 Q9LAT7	Q9LAT7 listeria mo
19	472	17.2	550	16 Q9CIL3	Q9CIL3 lactococcus
20	471.5	17.2	530	16 Q31315	Q31315 lactococcus
21	467.5	17.1	514	2 Q53480	Q53480 borrelia bu
22	467	17.1	545	16 Q9CIL2	Q9CIL2 lactococcus
23	446.5	16.3	553	16 Q927S4	Q927S4 listeria in
24	446	16.3	551	2 Q9F5U6	Q9F5U6 bacillus th
25	446	16.3	565	16 Q97D48	Q97D48 clostridium
26	441	16.1	529	2 Q9F1J4	Q9F1J4 enterococcu
27	440.5	16.1	519	2 Q9RMW8	Q9RMW8 bacillus an
28	431	15.7	543	16 Q9K114	Q9K114 vibrio chol
29	429.5	15.7	523	2 Q31313	Q31313 borrelia bu
30	429.5	15.7	529	2 Q31303	Q31303 borrelia bu
31	428.5	15.6	529	16 Q50927	Q50927 borrelia bu
32	428.5	15.6	530	16 Q985N6	Q985N6 rhizobium l
33	427.5	15.6	523	2 Q52615	Q52615 borrelia bu
34	425.5	15.5	529	16 Q51307	Q51307 borrelia bu
35	421	15.4	549	16 Q92FA6	Q92FA6 listeria in
36	419	15.3	426	16 Q84141	Q84141 chlamydia t
37	415.5	15.2	541	2 Q52616	Q52616 borrelia bu
38	414.5	15.1	541	16 Q51308	Q51308 borrelia bu
39	412	15.0	428	16 Q9PKP9	Q9PKP9 chlamydia m
40	412	15.0	549	16 Q9C127	Q9C127 lactococcus
41	411.5	15.0	523	2 Q31304	Q31304 borrelia bu
42	409.5	15.0	541	2 Q31306	Q31306 borrelia bu
43	400.5	14.6	547	16 Q83594	Q83594 treponema p
44	390.5	14.3	545	2 Q51643	Q51643 enterococcu
45	379	13.8	522	2 Q06515	Q06515 streptococcu

ALIGNMENTS

RESULT 1

Q928Y9 PRELIMINARY; PRT; 532 AA.

AC Q928Y9; 1000 (TREMBLrel. 10, Created)

DT Q928Y9; 1000 (TREMBLrel. 10, Last sequence update)

DT Q928Y9; 1000 (TREMBLrel. 19, Last annotation update)

DE OLIGOPEPTIDE BINDING PROTEIN.

GN OPPA1 OR CPN0195 OR CP0572.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029.

RX MEDLINE=99206606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,

RA Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

RT pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;

RX MEDLINE=20330349; PubMed=10871362;

RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,

RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;

RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RT from Japan and CWL029 from USA.
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL; AE001606; AAD18348.1;
 DR EMBL; AE002216; AAF38391.1;
 DR EMBL; AP002545; BAA98405.1;
 DR TIGR; CP0572;
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5.1.
 KW Complete proteome.
 SQ SEQUENCE 532 AA; 59744 MW; 1CB473D9D46A1579 CRC64;

Query Match 100.0%; Score 2739; DB 16; Length 532;
 Best Local Similarity 100.0%; Pred. No. 1.5e-196;
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKKISVGICITILLSVVLQCKESSHSTSRGELAINIRDEPRSLDPRQVRLLEISL 60
 Db 1 MKKISVGICITILLSVVLQCKESSHSTSRGELAINIRDEPRSLDPRQVRLLEISL 60
 QY 61 VKHIYEGVLQENNLGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFIESWK 120
 Db 61 VKHIYEGVLQENNLGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFIESWK 120
 QY 121 OVATQEVSGIYAFALPNKVKIKQEGHLSIDHFGVHSPNESTLVVLTLESPTSHFLKLLA 180
 Db 121 OVATQEVSGIYAFALPNKVKIKQEGHLSIDHFGVHSPNESTLVVLTLESPTSHFLKLLA 180
 QY 181 LPVFFVHKSQRTLSKSLPIASGAFTPKNIKQKOWIKLSKNPHYNSQVETKTIHF 240
 Db 181 LPVFFVHKSQRTLSKSLPIASGAFTPKNIKQKOWIKLSKNPHYNSQVETKTIHF 240
 QY 241 IPDNTAAKLFNQGKLNWQGPWGERIPQETLSNLSQKGLHSFVAGTSMLTENINKFP 300
 Db 241 IPDNTAAKLFNQGKLNWQGPWGERIPQETLSNLSQKGLHSFVAGTSMLTENINKFP 300
 QY 301 LNNMKLEALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERHQKQMAQRAYAKKL 360
 Db 301 LNNMKLEALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERHQKQMAQRAYAKKL 360
 QY 361 FKALELEQITAKDLEHLNLIFFVSSASSLLVQIREQWESLGFAIPVIGKEFALLOA 420
 Db 361 FKALELEQITAKDLEHLNLIFFVSSASSLLVQIREQWESLGFAIPVIGKEFALLOA 420
 QY 421 DLSSGNFSLATGWADPADMAFLTIFFAYPSGVPYPAIYNNKDFLEILQNEQEQDQHR 480
 Db 421 DLSSGNFSLATGWADPADMAFLTIFFAYPSGVPYPAIYNNKDFLEILQNEQEQDQHR 480
 QY 481 SELVSOASLYLETHPIPIYHDAFOFAMNKKLSNLGVSPTGVVDVFRYAKEN 532
 Db 481 SELVSOASLYLETHPIPIYHDAFOFAMNKKLSNLGVSPTGVVDVFRYAKEN 532

RESULT 2
 ID 084201 PRELIMINARY; PRT; 518 AA.
 AC 084201;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DE OLIGOPEPTIDE BINDING PROTEIN.
 GN OPPA_3 OR CT198.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=D/UN-3/CX;
 RX MEDLINE=9900803; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 "Genome sequence of an obligate intracellular pathogen of humans;

RT
 pnet
 Nuc.
 EME

RT Chlamydia trachomatis.
 RL Science 282:754-759(1998).
 DR EMBL; AE001293; AAC67790.1;
 DR InterPro; IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5.1.
 KW Complete proteome.
 SQ SEQUENCE 518 AA; 58858 MW; 8B3AE840831BBEF1 CRC64;
 Query Match 49.8%; Score 1364; DB 16; Length 518;
 Best Local Similarity 50.1%; Pred. No. 9.8e-94;
 Matches 269; Conservative 97; Mismatches 147; Indels 24; Gaps 8;
 QY 1 MKKISVGICITILLSVVLQCKESSHSTSRGELAINIRDEPRSLDPRQVRLLEISL 56
 Db 1 MKKISVGICITILLSVVLQCKESSHSTSRGELAINIRDEPRSLDPRQVRLLEISL 56
 QY 57 ELSLVKHIYEGVLQENNLGNIEPALAEDYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFI 116
 Db 57 DINLHLHYEGLVQETP-SGEVFPALAESFFLSDDKTYTFNLKKAFFWSNGDLTAEDFV 115
 QY 117 ESKQVATQEVSGIYAFALPNKVKIKQEGHLSIDHFGVHSPNESTLVVLTLESPTSHFL 176
 Db 117 ESKQVATQEVSGIYAFALPNKVKIKQEGHLSIDHFGVHSPNESTLVVLTLESPTSHFL 176
 QY 177 KLLALPVFFVH-KSORTLSQSKSLPIASGAFTPKNIKQKOWIKLSKNPHYNSQVETKT 235
 Db 177 KLLALPVFFVH-KSORTLSQSKSLPIASGAFTPKNIKQKOWIKLSKNPHYNSQVETKT 235
 QY 236 ITIHFIPDNTAAKLFNQGKLNWQGPWGERIPQETLSNLSQKGLHSFVAGTSMLTEN 295
 Db 236 ITIHFIPDNTAAKLFNQGKLNWQGPWGERIPQETLSNLSQKGLHSFVAGTSMLTEN 295
 QY 295 IKFPLNMMKLEALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERHQKQMAQRAYAKKL 355
 Db 295 IKFPLNMMKLEALASALDKEALVSTIFLGRKTAADHLLPTNHSYPERHQKQMAQRAYAKKL 355
 QY 355 YAKLFEALAELEQITAKDLEHLNLIFFVSSASSLLVQIREQWESLGFAIPVIGKEF 415
 Db 355 YAKLFEALAELEQITAKDLEHLNLIFFVSSASSLLVQIREQWESLGFAIPVIGKEF 415
 QY 415 ALQADLSGNFSLATGWADPADMAFLTIFFAYPSGVPYPAIYNNKDFLEILQNEQEQD 475
 Db 415 ALQADLSGNFSLATGWADPADMAFLTIFFAYPSGVPYPAIYNNKDFLEILQNEQEQD 475
 QY 475 DHKRSSELVSOASLYLETHPIPIYHDAFOFAMNKKLSNLGVSPTGVVDVFRYAKEN 532
 Db 475 DHKRSSELVSOASLYLETHPIPIYHDAFOFAMNKKLSNLGVSPTGVVDVFRYAKEN 532
 QY 532 NPKRSALISEASLVIERQNVIEPLYHDVPHYTNKLSFVFLHPSGLVDVRYAKNS 518
 Db 532 NPKRSALISEASLVIERQNVIEPLYHDVPHYTNKLSFVFLHPSGLVDVRYAKNS 518

RESULT 3
 ID 09PKJ4 PRELIMINARY; PRT; 520 AA.
 AC 09PKJ4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN,
 GN PUTATIVE.
 GN TC0471.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=MOPN / NIGG;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman K., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	40	2.2	914	4	US-09-177-234-2		Sequence 2, Appli
2	39	2.2	2019	2	US-08-243-511-46		Sequence 46, Appl
3	39	2.2	2019	2	US-08-600-993A-46		Sequence 46, Appl
4	37.2	2.1	1773	1	US-08-356-340-3		Sequence 3, Appli
5	37.2	2.1	1773	2	US-08-786-555-3		Sequence 3, Appli
6	35.8	2.0	7218	1	US-08-233-163-14		Sequence 14, Appl
c 7	35	1.9	1829	2	US-08-687-800-57		Sequence 57, Appl
8	35	1.9	3549	4	US-09-361-862-1		Sequence 1, Appli
c 9	34.8	1.9	8342	3	US-08-545-860B-63		Sequence 63, Appl
c 10	34.8	1.9	8342	5	PCT-US94-04496-63		Sequence 63, Appl
c 11	34.4	1.9	2340	4	US-09-450-852-3		Sequence 3, Appli
c 12	33.2	1.8	3350	3	US-08-617-860B-3		Sequence 3, Appli
c 13	32.8	1.8	307	1	US-08-241-943-5		Sequence 5, Appli
c 14	32.8	1.8	307	1	US-08-378-588-5		Sequence 5, Appli
c 15	32.8	1.8	307	2	US-08-811-094-5		Sequence 5, Appli
c 16	32.8	1.8	307	5	PCT-US94-11121-5		Sequence 5, Appli
c 17	32.8	1.8	2168	3	US-08-749-522-6		Sequence 6, Appli
c 18	32.8	1.8	2539	1	US-07-885-970A-28		Sequence 28, Appl
c 19	32.8	1.8	2539	1	US-08-238-687B-28		Sequence 28, Appl
c 20	32.8	1.8	2539	1	US-08-238-829-28		Sequence 28, Appl
21	32.8	1.8	5890	1	US-07-928-464-6		Sequence 6, Appli
22	32.8	1.8	5890	5	PCT-US93-07347-6		Sequence 6, Appli
23	32.8	1.8	6312	1	US-08-003-311B-6		Sequence 6, Appli
24	32.8	1.8	6312	1	US-08-261-432-6		Sequence 6, Appli
c 25	32	1.8	619	4	US-09-385-982-211		Sequence 211, App
c 26	32	1.8	1948	1	US-07-849-438-1		Sequence 1, Appli
c 27	32	1.8	4002	2	US-08-540-804-11		Sequence 11, Appl

ay 306 ttcaagaaaaataattttcaggaaatatagagctgctgttcgaagaactactctttt 365
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ab 152 TGCTAGAAATGATTCGCTACGGGCACTTTGTGCCGTCTTGGCTCAGATTGGTCTCAT 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||


```

; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgt-F1s
; US-08-232-463-14

```

```

Query Match          2.08; Score 35.8; DB 1; Length 7218;
Best Local Similarity 5.4%; Pred. No. 1;
Matches 22; Conservative 205; Mismatches 182; Indels 0; Gaps 0;

Qy 338 cctgcttgcagaagactctcttctcggcggaactcattatcttaaaactg 397
Db 1072 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1131

Qy 398 aaatcagcttttggagtaatggcgaccccttaacagctgaagactttatagaaatctgg 457
Db 1132 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1191

Qy 458 aaacaagtagactcaagaagctcagggaactctatgcttgccttggaatccaataaa 517
Db 1192 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1251

Qy 518 atgtacgaagatccaagaggacacctctcatagaccatttggagtggaactctct 577
Db 1252 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1311

Qy 578 atgaatcatcaactgttgaacctggaaatcccaacctgcatttcttaaaacttta 637
Db 1312 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1371

Qy 638 gctctcagcttttcccgcttcaaatcgaagaacctggaatccaatactctca 697
Db 1372 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1431

Qy 698 cctatagcaagcggagctttctctcaaaatatacaaaacaaacaaat 746
Db 1432 YYYGTAACAAATCTCTTACTCTTAACTACTGCTAGTAGGTAAT 1480

```

```

RESULT 7
US-08-687-080-57/c

```

```

; Sequence 57, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1829 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: INTRON 1 OF RAD50 GENOMIC SEQUENCE
; US-08-687-080-57

Query Match          1.9%; Score 35; DB 2; Length 1829;
Best Local Similarity 56.5%; Pred. No. 0.85;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 1477 caaccataaggactctcctagaaattctacaaaacatagacaagagcagatccaaaa 1536
Db 803 CAATCTAGTACATAACAAAGCATTCACAAATAACATTTTACAAAACAAAGTAAC 744

Qy 1537 acgtcgggaattagtgcgaagcttctctttacccatagagaccttctattatt 1591
Db 743 AAGTTTGGCAATCAATTCTCAACCTTCTCTCTTCTACACTCTTCACAAATCCTT 689

RESULT 8
US-09-381-862-1
; Sequence 1, Application US/09381862
; Patent No. 6245906
; GENERAL INFORMATION:
; APPLICANT: Ueyama, Hiroshi
; APPLICANT: Abe, Kanako
; APPLICANT: Keshi, Hiroyuki
; APPLICANT: Matsuhisa, Akio
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS
; TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

```

STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/381,862
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: JP 1997-71077
FILING DATE: 25-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP98/01288
FILING DATE: 23-MAR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Cawley, Jr., Thomas A.
REGISTRATION NUMBER: 40,944
REFERENCE/DOCKET NUMBER: 19036/36274
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3549 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: Clinical Isolate Sp-6-28
US-09-381-862-1

Query Match 1.9%; Score 35; DB 4; Length 3549;

Best Local Similarity 61.5%; Pred. NO. 1.2;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY	306	ttcagaagaataattcttcaggaaatatagacctctctgcagagactactctctt	365
Db	2486	TTTGGAAACAGTGACTATGGTAAATTAGTACCATCATTCTCTAAAGATTGGAAGGTT	2545
QY	366	cctcgacgcgactcactatactttaact	396
Db	2546	CAAAAGCGTCTGACTTATACATTACTCT	2576

RESULT

US-08-545-860D-63/c
Sequence 63, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaanani, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESSEE: No. 6040140ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:

Query Match

1.9%; Score 34.8; DB 3; Length 8342;


```

; NAME/KEY: CDS
; LOCATION: 3032..3145
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6788..6934
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7967..8062
; NAME/KEY: CDS
; LOCATION: 8304..8342
;
PCT-US94-04496-63

Query Match
Best Local Similarity 1.9%; Score 34.8; DB 5; Length 8342;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 658 cgttcataaatctcaagaaccctgcgaatccaaatctctacctatagcaagcggagcttt 717
Db 3302 CCTTCTTAATCTCTAATACTTCAAAACATTTCTTAGTGTGTTTAAACAGGGAATA 3243

QY 718 ctatcctaaaaatatacaaaaaaacaatggataaaactctcaaaaaacccctcactacta 777
Db 3242 GTACCTGTAAATACCAACAAACAGAGATAGTAATCTCATCCCGAGATCAATCATTTA 3183

QY 778 taatcaagtcaggtggaaactaaacacattacgattcacttcattcccgatgcaaacac 837
Db 3182 TGTACTTCTCGGGGTACAAAGACGAGATGCCTTACCTCTACATGCCCTACTGTGGCAC 3123

QY 838 ag 839
Db 3122 AG 3121

RESULT 11
US-09-450-852-3
; Sequence 3, Application US/09450852
; Patent No. 6309860
; GENERAL INFORMATION:
; APPLICANT: Chaganti, Raju S.K.
; APPLICANT: Dyomin, Vadim
; APPLICANT: Dalla-Favera, Riccardo
; TITLE OF INVENTION: CLONING AND USES OF BCL-8
; FILE REFERENCE: 53828-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/450,852
; CURRENT FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: human
US-09-450-852-3

Query Match
Best Local Similarity 1.9%; Score 34.4; DB 4; Length 2340;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1668 ttgtggacttcggttatgctaaggaaaattagcaactcttttaactcgcgaactgtca 1727
Db 1970 ttattgactacactattctgaggctactggcaccctctctgatttcatactattgtaa 2029

QY 1728 agaactgaatcttataactaaactgggtgctt 1759
Db 2030 tgacattagatatttcttatattgtctgcatt 2061

RESULT 12
US-08-617-860B-3/c
; Sequence 3, Application US/08617860B
; Patent No. 6133506

```

```

;
; GENERAL INFORMATION:
; APPLICANT: Tvper, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hvrice-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Miller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,860B
; FILING DATE: 01-MAR-1996
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02950
; FILING DATE: 05-SEP-1994
; APPLICATION NUMBER: DE P4329951.2
; FILING DATE: 04-SEP-1993
;
; INFORMATION FOR SEQ ID NO: 3:
; .SEQUENCE CHARACTERISTICS:
; LENGTH: 3350 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; IMMEDIATE SOURCE:
; LIBRARY: genomic Lambda FIX II
; CLONE: BnRCaseg10
; FEATURE:
; NAME/KEY: Startcodon
; LOCATION: 2611..2613
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2611..2908, 3001..3341)
;
US-08-617-860B-3

```

```

Query Match      1.8%; Score 33.2; DB 3; Length 3350;
Best Local Similarity 56.4%; Pred. No. 4.3;
Matches 62; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1131 aacatcaaaacgaagatggcacaacgccaagcttacgtcaaaaactctttaaagaag 1190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1999 AACTTTAAATACAAATAAACCAAAATAAAGATCATTTGTAATAAATGCAATAACAG 1940
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1191 ctttagaagaactccaatcaactgctaagaagatctcgaaacatcttaattctt 1240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1939 AAATAAATAACACCAAAATCACTTCAACTAATTTGGTTCTCTTACCAT 1890
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-08-241-943-5/C
; Sequence 5, Application US/08241943
; Patent No. 5602321
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS BIOPLASTIC
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady

```

```

;
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: First Wisconsin Plaza, One South
; STREET: Pinckney St.,
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/241,943
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: us/07/980,521
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 11-229-9076-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-2484
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 5:
; .SEQUENCE CHARACTERISTICS:
; LENGTH: 307 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium barbadense
; STRAIN: Sea Island
; IMMEDIATE SOURCE:
; LIBRARY: EMBL-SI
; CLONE: B8
;
US-08-241-943-5

```

```

Query Match      1.8%; Score 32.8; DB 1; Length 307;
Best Local Similarity 53.0%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 627 taaaacttttagctctccagctctttcccggttcataaaatctcaagaaccctgcaat 686
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 TGAACACTTTTAGTTTTCGAATAATTTTACCGCGATCTATTTCGGTAATGATGTAAT 155
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 687 ccaatctctaccatagcaagcgagctttctatctctcaaaaataatcaaaaaaacaaat 746
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 AAAAAATTTTCGATGGTTTTTAAATAAGATATCGGATTCAAATATATATAAAAAATAATTT 95
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 747 ggaataaaactct 758
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 AAAAAAATTACT 83
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
US-08-378-588-5/C
; Sequence 5, Application US/08378588
; Patent No. 5608148
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady

```

STATE: WI
COUNTRY: USA
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,094
FILING DATE: 03-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/378,588
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 11-229-9101-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-2484
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium barbadense
STRAIN: Sea Island
IMMEDIATE SOURCE:
LIBRARY: EMBL-SI
CLONE: STB8
DS-08-811-094-5

Query Match 1.8%; Score 32.8; DB 2; Length 307;
 Best Local Similarity 53.0%; Pred. NO. 1.5;
 Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY	627	taaacatttagctcttcaglottttccccgttcaaatc	caagaaccctgcaat	686
Db	214	TCAACCTTTTACGTTTCGAAATATTTTACCCGATCTATTTCGGTATGATTGAGTAAT	155	
QY	687	tcaaatctctacctatagcaagcggagctttctatcctaa	aaatcaacaacaaacaaat	746
Db	154	AAAAATTCATGGTTTTTAAATAAGATATCGGATCAAAATATTATAAAAAATAATATT	95	
QY	747	ggataaaactct	758	
Db	94	AAAAAAATTAAT	83	

Search completed: July 26, 2002, 03:36:33
Job time: 2316 sec

[illegible]

```

QY 728 aatatcaaaacaaacaaatggataaaactctcaaaaacccctcaactactataatcaact 787
Db 457 MAMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 516
QY 788 cagggtggaacaaacaaatgattcagattcacttccattccatccatgatacaaacaca 838
Db 517 MAMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 567

RESULT 7
CNS0100X/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL098379
VERSION
AL098379.1 GI:5609990
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
Source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03G04"
/note="end : SP6"
BASE COUNT 195 a 108 c 131 g 161 t 506 others
ORIGIN

Query Match 2.6%; Score 46; DB 12; Length 1101;
Best Local Similarity 16.7%; Pred. No. 4;
Matches 117; Conservative 283; Mismatches 284; Indels 15; Gaps 2;

QY 515 aaaaatgtacgaagatccaaagaggacacctctccatagaccatttggagtgacct 574
Db 1101 MNDMVTMMHNNMTAMWTATGATGWKWTATMMHRTHTTHHHHTHHHHHTWYWTATTH 1042
QY 575 cctaatactacactgttggtaacctccatccatcccaactcgcattcttctaaact 634
Db 1041 KMMWTYRCYKCHYNNYNTCTHCTCCYCHCCMCWCWHAAMMMHMMHFWAWWTHT 982
QY 635 ttagtcttcagttttttcccgcttataatctcaaaacccctgcgaatcaa-alc 593
Db 981 WTMMWMMCTCNCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 922
QY 694 tctacctagcagcaggagctttctctctcaaaatatacaaaaacaaatggataaa 753
Db 921 WCHCTHMMHCTMCAHCHCTMAYMYMYMYMYMYCYCYCYCYCYCYCYCYCYCYCYCYCY 862
QY 754 actctcaaaaacccctcactactataatcaagtcagtggaataacaaacgattacgat 813
Db 861 CCYCYVNHMTATTCATATATATATATATATATATATATATATATATATATATATATAT 802

```

```

QY 814 tcatttcattccgatgcaaacagacagcaaaactatttaacaggaacaaactcaattg 873
Db 801 AHHHHYCTWTHHHTHHHCHHCHHTWTHHHHHYHTYTAAMMAAAAAAAMAAAAA 742
QY 874 gcaagacacctctggggaagaacgattctctcaagaacccctcatcccaattcagctaa 933
Db 741 AAAM-----MAAAMATWATATTTATTTATTTATTTATTTATTTATTTATTTATTT 596
QY 934 ggggcaactcaactcttttgatgcaggaacccctcagtcgctcaccctcctcaataaa 993
Db 695 HCHCCACCHMYHCHYCHMYHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMM 636
QY 994 attccctccaaataatgaagcttagaagccttagcatcacccttagataaagaagc 1053
Db 635 WCYMYCTMAHMYHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMM 576
QY 1054 tctgtctcaactatattcttgccgctgcaaaaactccgcatctctcctcactacaaa 1113
Db 575 YHTYTCHAYHTYHCHWCWCHAMYKYHAMMMACAYAMAMMAHAMAYMMMTMAMHMMHMM 516
QY 1114 tattcatagctatcccaacatcaaaaaacagagatggcacaacgcgcagcttagctaa 1173
Db 515 THTTWEHAWAHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMH 456
QY 1174 aaaaactcttaagaagcttttagcagaactcccaaatcac 1212
Db 455 HAWHHAHTMCWHAHAMHAWHHAHAMHAWHAWHAWHAWHAWHAWHAWHAWHAWHAWHAW 417

RESULT 8
CNS016JY
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16D23 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL106840
VERSION
AL106840.1 GI:5624081
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
Source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN16D23"
/note="end : T7"
BASE COUNT 337 a 71 c 48 g 61 t 584 others
ORIGIN

Query Match 2.58%; Score 45.6; DB 12; Length 1101;
Best Local Similarity 21.4%; Pred. No. 5;
Matches 82; Conservative 128; Mismatches 173; Indels 0; Gaps 0;

```



```

Query Match          2.5%; Score 45.2; DB 12; Length 859;
Best Local Similarity 20.6%; Pred. No. 6;
Matches 54; Conservative 104; Mismatches 104; Indels 0; Gaps 0;

Qy 611 ccaacctgcattcttctaaacttttagctctctcagctcttttccctctcataaattc 670
      | : | | | | : : : | : | | | : | : | : | : | : | : | : | : | : | : | : |
Db 596 CWHAWCTCWCYAYWMAAACCYATMCCACYACYYCCYCYMYCYCYYYTAAATYAY 655
      : | | | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 671 caaagcaacctgcgaatccaaactctacattctacatagcaagcgagctcttctatcctaataat 730
      : | | | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 HAAAYHVCYAAAHYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 715
      : | | | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 731 atcaaacacaaacaaatgataaaactctcaaaacccctcaactataataatcaaatcg 790
      : | | | : | : | : | : | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 716 YYAYHYVYVYAYVYAHYVYAHYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 835
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 791 gtggaactaaacagattacattcaactctcattcccgatcgcaaacacagcagcaaaaata 850
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 776 AAYAYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 835
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 851 tttaatcagggaacaaactcaatt 872
      : | | | : | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 836 HWYHFAAAAAAAAHAAHVYV 857
      : | | | : | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
CNS0168G/c 1201 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BAC15114 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106426.1 GI:5621843
VERSION AL106426
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1201)
          Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
          collaboration with the European Drosophila Genome Project (EDGP) -
          http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
          library (Dros BAC) was made by Alain Billaut at CEPH (Centre
          d'Etude du Polymorphisme Humain) with funding provided by a MRC
          project grant. The DNA was prepared from embryos by Alain Bucheton
          and Genevieve Payan. It has been constructed in the vector
          pBelobAC11.

FEATURES
          source
          1..1201
              /organism="Drosophila melanogaster"
              /plasmid="pBelobAC11"
              /db_xref="taxon:7227"
              /clone_lib="DrosBAC"
              /clone="BAC15114"
              /note="end : T7"

BASE COUNT 337 a 214 c 258 g 159 t 233 others
ORIGIN

```

[illegible]

RESULT 12	
CNS00FOO/c	
LOCUS	CNS00FOO 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR31F03 of APC-98 library from <i>Drosophila melanogaster</i> (fruit fly), genomic survey sequence.
ACCESSION	AL070854
VERSION	AL070854.1 GI:4950896
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	<i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> .
REFERENCE	1 (bases 1 to 1101)

COMMENT

web : www.genoscope.cns.fr/
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RP01-98 and was constructed by partial *NotI* digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

FEATURES	source
found at http://bacpac.med.buffalo.edu/drosophila	1. 1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone.lib="RPCI-98"
	/clone="BACR31F03"
	/note="end : 73"
BASE COUNT	167 a 84 c 86 g 155 t 609 others
ORIGIN	

Query Match 2.5%; Score 44.2; DB 12; Length 1101;

BASE COUNT

BASE COUNT 185 a 132 c 59 g 194 t
 ORIGIN

Query Match 2.4%; Score 43.8; DB 10; Length 570;
 Best Local Similarity 51.8%; Pred. No. 12;

Matches 99; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 599 accctggaatccccacccctgcattttttaaacttttagcttttccagctttttccccc 658
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 145 AACCTGGTACTCTTAACATACATATTTCTCAACATCTTTACACCTCTTTACACCTCTTC 204
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 659 gtccataatctcaaaagaccctgcatttcaaatctctacattatagcaagcgagctttc 718
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 205 CTCTTACTACCTCCACAGTAATTAATAACAAACAAACCTACTATAAAGAACTCTA 264
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 719 tatcttaaaatatacaaaaaaacatggataaaactctcaaaaaacccctcactactat 778
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 265 GATCCAAGATATTTTCATGCTTTAATCATGAAGTTTAATCACAAAAAATCTCCATTACCTT 324
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 779 aatcaaatgca 789
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 325 CATCTCTCTTCA 335

RESULT 15

CNS0073W/c

LOCUS

DEFINITION

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Determination of this BAC-end

collaboration with the Berkeley

The BDCP is constructing a physical

melanogaster genome using these

please see <http://www.fruitfly.org>

melanogaster BAC library was

Aaron Mamooser in Pieter de Jong's

NY. The library is named RPCI-98

isogenic strain y2; cn bw sp, the

Pl and EST libraries. A more detailed

and how to order individual BAC

filters for hybridization from the

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm

Location/Qualifiers

1..922

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR14D09"

/note="end : TET3"

BASE COUNT 223 a 95 c 109 g 221 t 274 others

ORIGIN

Query Match 2.4%; Score 43.8; DB 12; Length 922;

Best Local Similarity 23.0%; Pred. No. 13;

Matches 48; Conservative 77; Mismatches 84; Indels 0; Gaps 0;

QY 662 cataaatctcaagaacctgcatttcaaatctctacattatagcaagcgagctttctat 721
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 806 MAMMVAAC 747
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 722 cctaaaaatatcaaaaaaacatggataaaactctcaaaaaaacctcactactataat 781
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 746 CMMAAAMAMMMAACAMMAAAMMAAAMMAAAMMAAAMMAAAMMAAAMMAAAMMAA 687
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 782 caaagtccagtggaactaaacacgattacgattcacttcttcctccgattgcacacagca 841
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 686 CM 627
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 842 gcaaaactatttaatacagggaaaactcaa 870
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 626 AMMAAMMAAC 598
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: July 26, 2002, 03:34:28
 Job time: 2241 sec


```
/protein_id="CAD10302.1"
/db_xref="GI:16541563"
/translation="MRKISVGCITILLISLVVLOGCKESSHSRSBELAINIRDEP
RSLDPROVRLLSISLVKHYISGLIQENNLNSGNEPALAEDYISLSDGLTYTFKLKSA
PWSNGDLTADIESWKOVATQEVSGIYAFALNPINKVRKIQEGLSIDHFGVHSPN
ESTLVLTLESPTSHFLKLLALPVFFVHKQSRTLSKSLP IASGAFYPRNKIQKOWIK
LSKPHYINQSVETKTIITHEIPDANTAFLKLNQKINQWGPWGERIPDPTLSNLIQ
SKGHLHSFDVAGTWSLTFNINKFPLNNMKREALASALDKREALVSTIFLGRKTADEL
LPTNHSYPERHOQMAORQAVAKLFWKELEELQITAKDLEHLNLIPEVSSASSLL
VOLIREWKESLSFALPIVKGFSALLQADLSGNFSLATGCGNFWAFADPMFLFIYFAY
PSGVPVPAIHKHOFLEILQNIQEBQDHQKRSLSVQSLSLYLETFHFIIEPIYHDAFQFA
MKNKLSLNGVSPSTGVVDFRYAKEN"
BASE COUNT      560 a 439 c 294 g 506 t
ORIGIN

Query Match      100.0%; Score 1799; DB 6; Length 1799;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acttccccctgtcaaaactatgctcagataatgctgctatgattgacaggtctagggggag 60
DB 1 ACTTCCCCCTGCTAAACTATGCTCAGATAATGCTGCTATGATTCAGGCTTAGGGGGAG 60

QY 61 aaaaatttcaaaaaaactctagtattccggaaattogtatgogcaagatatcagtggg 120
DB 61 AAAAATTTCAAAAAAACTCTAGTATTCGGGAAATTCGTATATGCGCAAGATATCAGTGGG 120

QY 121 aatctgtatcacacctctctctagctctccgttagtctcccaagctgcaagagaccag 180
DB 121 AATCTGTATCACCATTCCTTAGCTCTCCGTAGTCTCCGAGCTGCCAAGGAGTCCAG 180

QY 181 tcaactctcaactctogggagaactcgctatttaataagaagatgaaccccgctttct 240
DB 181 TCACCTCCTACATCTCGGGAGAACTCGCTATTAAATAAGAGATGAACCCCGCTTTCT 240

QY 241 agatccaagacaagtggaactcttccaaataacagctcttccaaatactatcaggg 300
DB 241 AGATCCAGACAAAGTGGGACTCTTTCAGAAATCAGCCCTGTCAACATATCTATCAGGG 300

QY 301 attagtcgaagaaataatcttccagaaataatagacctgctctctcagaagactactc 360
DB 301 ATTAGTTCAGAAATAATCTTTTCAGAAATATAGAGCCTGCTCTTCACAGAACTACTC 360

QY 361 tcttctcctggaggaactcactatctttaaactgaaatcagctcttttggagtaatgg 420
DB 361 TCTTCTTCGGAGGAGTCACTTATACITTTTAAACTGAAATCAGCTTTTGGAGTAAATGG 420

QY 421 cgacctcttaacagctgaagactttatagaactcttggaaacaagtaactactcaagaagt 480
DB 421 CGACCCCTTAACAGCTGAAGACTTTATAGAACTCTTGGAAACAAGTACTACTCAAGAA 480

QY 481 ctcaagaatctatgcttttgccttgaaatcccaattaaaaatgtacgaagaatcaagagg 540
DB 481 CTCAGAAATCTATGCTTTTGCCTTGAATCCCAATTAATAATGTACGAAGATCCAGAGG 540

QY 541 acacctctccatagaacattttggagtgcaactctcctaataagaatctacacttggttac 600
DB 541 ACACCTCTCCATAGACCAATTTTGGAGTGCACTCTCCTAATGAATCTACACTTGTGTAC 600

QY 601 cctggaatcccccaacctcgattctctaaaccttttagctcttccagctcttttccccgt 660
DB 601 CTTGGAATCCCCACCTCGGATTCCTTAAACCTTTAGCTCTTCCAGTCTTTTCCCCGT 660

QY 661 tcataaatctcaagaagacctgcaatccaaatctctacatagcagcgaggtttctcta 720
DB 661 TCATAAATCTCAAGAAGACCTGCAATCCAAATCTCTACCTATAGCAAGCGGAGCTTTCTA 720

QY 721 tctcaaaaatacaaaaacaaatcggataaaaactctcaaaaaacccctcactactataa 780
DB 721 TCTTAAAAATATCAACCAAAACAAATCGGATAAAATCTCAAAAACCCCTCACTACTATAA 780

QY 781 tcaagtcaggtgggaactaaaacagattacagttacattcattcccgatgcgaacacagc 840
```

RESULT 2
AE001606
LOCUS
DEFINITION
Chlamydia pneumoniae section 22 of 103 of the complete genome.
ACCESSION AE001606 AE001363

781 TCARAAGTCAGTGGAAACTAAACAGATTACGATTCACTTCATCCCGATGCAACACAGC 840
QY 841 agcaaaactatttaatacagggaaaaactcaattggcaaggacctctcttggggagaaacgat 900
DB 841 AGCAAAACTATTTAATCAGGAAAACTCAATTGCACTTAAGGGGCACTTACACTCTTTTGTATGTCG 900
QY 901 tctcaagaaaacctctatccaaattacagctcaagggaacctacactcttttgatgtcgc 960
DB 901 TCCTCAAGAAACCTTATCCAATTACAGCTTAAGGGGCACTTACACTCTTTTGTATGTCG 960
QY 961 aggaacctctatgctcacctccaattcaataataatccccctcaacaatatgaagcttag 1020
DB 961 AGGAACCTCATGCTCAGCTCAGCTCAATATCAATTAATTTCCCTCACAATATGAAGCTTAG 1020
QY 1021 agaagccttagcatcagctcagcttagaagaagacctcttgcctcaactatattcttaggcg 1080
DB 1021 AGAAGCCTTAGCATCAGCCTTAGATAAGGAAGCTCTTGTCTCACTATATATTCTTAGGCGC 1080
QY 1081 tgcataaaactgcgatactctcctcaactataaaattattcagctatccccgaacataaaa 1140
DB 1081 TGCATAAACTGCGGATCATCTCTTACCTTACAAAATATTTCATAGCTATCCCGCAACATCAAAA 1140
QY 1141 acaagagatggcacacgccaagcttacgcttaaaaaactctttaagaagcttttagaaga 1200
DB 1141 ACAAGATGCGCACACGCCCAAGCTTAGCTTAAAGAACTCTTTTAAAGAAAGCTTTAGAAGA 1200
QY 1201 actccaaactcagctgaagaatctcgaacatcttaattcttcttcccgcttctcgc 1260
DB 1201 ACTCCAAATCAGCTGCTTAAAGATCTCGAATCTTAATCTTATCTTTTCCGTTTCTCTGTC 1260
QY 1261 agcaagctcttactagtcacaaacttatcagacacagctggaagaagcttttaggttcgc 1320
DB 1261 AGCAAGCTCTTACTAGTCCAACTTATACGAGACAGTGGAAAGAAAGTTTAGGGTTCG 1320
QY 1321 tatccctattgtcgaaaggaaatttgcctcttcccaagcagacctatcttcagggaact 1380
DB 1321 TATCCCTATTGTGCGAAAGAAATTTGCTCTTCTTCCAAAGCAGACCTATCTTCAGGAACTT 1380
QY 1381 ctcttagctacagaggagatggttcgacagatttgcgtactatggcatttctaagcgt 1440
DB 1381 CTCCTTAGCTACAGGAGGATGGTTGCGAGACTTTGCTGATCTTATGSCATTTCTTAACGAT 1440
QY 1441 ctttgccttaccatcagaggttctctctctatgcaatcaaccataagagacttccataaat 1500
DB 1441 CTTTGCTTATCCATCAGAGTTCCTCTCTTATGCAATCAACCATAGGACTTCTCTAGAAAT 1500
QY 1501 tctcaaaaacatagacaagagcaagatcaccataaacgctcggaattagtgctgcaagc 1560
DB 1501 TCTACAAAACATAGAACAAAGAGCAAGATCACCATAAACGCTCGGAATTAGTGTGCGCAAGC 1560
QY 1561 tctcttaccctagagaccttctctattattgacgcgacttaccacagcagcattcaatt 1620
DB 1561 TTCTCTTTACCTTAGAGACTTTCATATATTATTGAGCGGATCTACCACGACGATTTCAATT 1620
QY 1621 tgcctgaataaaaaaactttctaatctaggagctcaccacaacagaggttgtagactcgc 1680
DB 1621 TGCTATGAATATAAATCTTCTATCTTAGAGTCTCACCAACAGGAGTGTGGACTTCG 1680
QY 1681 ttatgctaaaggaaaattagcacctcttttaattctcgaaaacttgcgaagaactgaactt 1740
DB 1681 TTATGCTAAGGAAAAATTAGCACCTCTTTTAACTCGCAAACTTGTCAAGAACTGATCTT 1740
QY 1741 atactaaactgggtgcctttgtggacctcgtttctctctctgactgctctctctctcta 1799
DB 1741 ATACTAAACTGGGTGCCTTTGTGSCACCTCGTTCCTTCTGACTGCTCTTCTCTCTA 1799

```

VERSION
KEYWORDS
SOURCE
ORCANSIM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS

AB001606.1 GI:4376464
Chlamydomophila pneumoniae CWL029.
Chlamydomophila pneumoniae CWL029.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
1 (bases 1 to 11648)
Kallman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,
Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)
99206606
10192388
2 (bases 1 to 11648)
Kallman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R. W. and Stephens, R. S.
Direct Submission
Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
Location/Qualifiers
1..11648
/organism="Chlamydomophila pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
73..1107
/gene="gcp_2"
/feature="gcp_2"
73..1107
/gene="gcp_2"
/codon_start=1
/transl_table=11
/product="O-Sialoglycoprotein Endopeptidase"
/protein_id="AAD18347.1"
/db_xref="GI:4376465"
/translation="MLTIGLSSGDEYACALVNEQKQILANIASQDIHASYGVVPE
LASRLHHPVQVINKALCOANLIEDMLIAVTPGLISLGVGHFGKGIAGAK
ASLGVNVEAHLYAYMAQNVQFPALGVVSGAHTAAFFIENPYSKLLGKTRDA
IGFTDKVGRFLGYPAGPLIEKIALGESDSEYSPKAPVNDYFSGSLKTAVALY
IKGNSSPAPAEISLEKQDIAAFKAACTTIAOKLPTIIEFSCRSILIGGVA
INEYFRSAIQTACNLPPVFPAPKLCSDNAMIAGLGENFQNKSSIPERICARYQE
SVSPFSLASP"
1033..2651
/gene="oppA_1"
/feature="oppA_1"
1053..2651
/gene="oppA_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAD18348.1"
/db_xref="GI:4376466"
/translation="MRKISVGICITILLSVVLQCKESSHSTSRGELAINTRDEP
RSLDPRVRLSEISLVKHIYEGVAVQENNLGNIEPALAEYSLSLSCGLIYTKLSA
FWSNGDPLTAEDFIESMKQVATQVSGIYAFALNPKVNRKIQGHLSIDHFGVHSPN
ESTLVYIESPSHSLKLLALPFEFVHKRSORTLOSKSLPIASGAFYKNIKQKOWIK
LSKNPHYQSQSVETKLTITIHFPDANTAKFNQCKLNMWQGPWGRIPQETLSNLIQ
SKGHSFVAGTSVLTFTENKFNPLNNKRLREALASLDKEALYSTIFLGKRTADHL
LPTNHSYEPHQKQMAQRAYAKKLFKALPELOITAKDLEHLNLPFVSSASSILL
VOLIRBOWKESGFAIPVTKERALLQADLSGNSRLAIGWGFADPADPWFATIFAY
PSGVPIYANKHDKLEITLQNLREQDQKRSSELVSQASLYLETFHIEPIYHDAFOFA
MKKLSNGVSPGTGVDFRAKEN"
2802..4385
/gene="oppA_2"
/feature="oppA_2"
2802..4385
/gene="oppA_2"
/codon_start=1
/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAD18349.1"
/db_xref="GI:4376467"
/translation="MLRFAVFIETLWLTISGCSQSSQSGKGFVYNNKEMPRSLDFGK
TRLADQIMRLHYELGVEHSQNGIEPALAEYSITSEDTGYTEFKIKNLWNSGDP
LTAQDFVSSWEIKLEDASSVLYAFPIKNARAIFDDTESPGLVRAIDKRLHLEIQ
LETPCAHLHFTLPIFFPVHETLRNYSTSFSEMPITCGAFRPVSLKGLRLHLEKNP
MYHNSRVKLIIVQFISNANTAALFKKKLDWQGPWPGEPIPPISASLHQDQL
FSLPGASTWLLFNIOKQPNNAKALKSLAIDKMLTKVYOGLAEPDHLHPLRL
YPGTYPERKRONREILAQOLFBEALDELOMTREDEKETITFTSFSESYGRICOMLR
EOWKVKLPTIPVQBEFTIKNFLEGNYSLTWNQWTAARIPKMSYLMIFANPGGIS
PYHLODSFQVILLIKIQEHKKHLRNQLIIEALDYLEHCHLEPLCHPNLRIALNKN
KNFNLFVRSIDSRFIEKL"
4441..5748
/gene="oppA_3"
/feature="oppA_3"
4441..5748
/gene="oppA_3"
/codon_start=1
/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAD18350.1"
/db_xref="GI:4376468"
/translation="MFSRWITLLELFIISLTGCSYSSKHKOSLIIPIDHPVAFSPEQ
AKRAMDLSIAQLLEFGLTRHRESDLELAISRYTSEDFCSYTFIFKDSALMSDG
TPTISEDIRNAYEAQENSPIQIFQGLNFTSPSNATITHLDSNPDPKLLAFPAF
AIFKPNKFLSGPYTLVEYFPGHNIHLKKNPYDYHCVINSIKLLIIPDIYAIH
LLNRKGVDMWGPWHQGIPEWLHKQSOHYTYTVEGAFWCLNTKSPHLNDLONHR
LATCIDKRSIIIEALQGTQOPATLSRGAPQPNQYKQKPLTQEKLVITYPSDLRC
QRIASILEQWKAAGIDILLEGLEYHLFVNRKKVODYATATGTGVAYYFGANLISED
KLLQNFELIPIYILSYDYILFQDFIEGVINAGAVDLKYTFP"
6032..7612
/gene="oppA_4"
/feature="oppA_4"
6032..7612
/gene="oppA_4"
/codon_start=1
/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAD18351.1"
/db_xref="GI:4376469"
/translation="MHRKAPTLKSLIPNLFLITLSSCKQKQEPGLKHLVIMASHD
LADLNPNAVYLSASLAKALYEGUTRETDQGIALAEYSITLSKDKHYTFKLRPSV
WSDGTPTAYDFEKSILQYIEEFSPIHTLLGVKNSSAIHNAOKSLETLGIAOKD
LTVTIIEQPPFPFLTIARPVFSPVHHTLRESYKKGTPPSTVINSNGPVLKKEHON
YLILEKNPHYDHESVKLDRLVKLITIPDASTATKLEKSIDMIGSPWSPAPISNEDQK
VLSQEKULTYVSSTLLIYNQPLQNKALRAIAHADRKSILRLVPSCQEAATL
VPPNLSQNLQKISITEERQTKARAYFQEKARELSEKELAEISILYPIDSSNSITIAQ
EIQRKLDLGLKIKIQGMEYHCFILKRRQGDFFIATGGWIAEYVSPVAFSLILGNP
DLTQWRNSDYETLEKLYLPHAYKENLRAEMIEEETPIIPLYHGKYIYAIHPKION
TFGSLGHTDLKNIDILS"
7905..8849
/gene="oppA_1"
/feature="oppA_1"
7905..8849
/gene="oppA_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAD18352.1"
/db_xref="GI:4376470"
/translation="MFSYIKRILFNLLSLWILITLFLVMKTIQDGFENDEGNVLIS
BEVLOTLSRYGDKPLVQOYTOYLSAKLDGNSLWYKDRVYNIISTAPFISAIL
GLQSLFISGGTALGTIAALKKKORRYILGASILQISIPAFIATILQYVFAKLP
LIPACGWSFTHTIPLTALAVTPMAFIQLTYTSSSAALNKDYVLLAYAKGLSPKV
VILHYPYALPFTISYSFALTTIVITGFALENIFCIPGLGKWFICISIKQRDYPVALG
LSVFYGLFMLSLSLTIQSIIDPQIRYAHGKEKKR"
8880..9734
/gene="oppC_1"
/feature="oppC_1"
8880..9734
/gene="oppC_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAD18353.1"
/db_xref="GI:4376471"
/translation="MENLSSAPSSIWKSIIQNKMLVIGLTLTILMLGALLPWFYQ
DYQTSIKDILVSPCSRPFFGTGLGRCMFARTLRLRLSLIATLTIATLIDVCGVGLM

```


ATVAGGKKIDFLMRTELEIPRIPIIILLIVIFHHGLLPLILAMITITGWIPIS
RIYGFLLKKNKPVLSAKAMHASTPHILKKNHLLPLAPIISTLIPTIPNAYTEA
FISFGLGQPQASGLVKEGINAIDYPLWFFPPLSLIMIALSISFNLIIGEGAKTL
CLEEGSHG"

gene

9727..10581

CDS

/gene="oppb"

/note="Cp0201"

9727..10581

/gene="oppb"

/codon_start=1

/transl_table=11

/product="Oligopeptide Transport ATPase"

/protein_id="AAB18354.1"

/db_xref="GI:4376472"

/translation="MDNYLLNIDKLTISTNPKRTLIENLSLQKRNRLALVSGSGS
GKTTIKALIGFLPENCILKTSILFEDIDITKLSPELKHKIRGQKIATILQAMGSL
TPSMRIGMIIETLRHKKNKKEAYNKMOLLTDVCPNPKPFSFYSPFELSGMRQ
RVIAIALASPKLILADEPTALDSMSOAVLRILRNIOOQOATILAVTHNLSLVK
ELNDICILKDGKLIETGTVEEIFLSPKHPVTLKLLNAVSKIPIKTSSPILKNKFP
LMSMGGL"

gene

10578..11366

CDS

/gene="oppb"

/note="Cp0202"

10578..11366

/gene="oppb"

/codon_start=1

/transl_table=11

/product="Oligopeptide Transport ATPase"

/protein_id="AAB18355.1"

/db_xref="GI:4376473"

/translation="MTLLSIKDLSTIRGKKILNHLINLIKSYLTIVPSPSGKS
SLAUIIDLLKPTTITFHMDPKIPRAKRVQVIVDDISLNPQMSIKGIISPLNI
IGTYSKAEQNKETIYNVLDVNLPSVLHKLPSGQKORIALAKALVSKPELLICD
BPLSLDTLMSLILDLFOTIKKEONTLPFTHDSAAVYIADIIAMVQDSLVERH
CREKIFTPKHITTDLLDAIDIESLISTEMEPSEYELQVASK"

BASE COUNT

3743 a 2799 c 1831 g 3275 t

ORIGIN

Query Match 100.0%; Score 1799; DB 1; Length 11648;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acttccccctgctaaactatgctcagataatgctgctgctatgctgaggtctagggggag 60
|||||
DB 953 ACTTCCCCCTGCTAAACTATGCTCAGATAAATGCTGCTATGATTCGAGGCTCAGGGGAG 1012
|||||

QY 61 aaattttcaaaaaactctagatttccgggaattcgtatgctgcgaagatatcagtg 120
|||||

DB 1013 AANAATTTCAAAAAAATCTAGTATTCCGGAAATTCGTATATGCGCAAGATATCAGTGG 1072
|||||

QY 121 aatctgtatcaccattctccttagcctctccgtagctcctcgaagctgcaagagtcag 180
|||||

DB 1073 AATCTGTATCACCATTCTCCTTAGCCTCCTCCTAGCTCCTCAGGCTGCAAGAGTCCAG 1132
|||||

QY 181 tcaactcctcatcctcgtgggagaactcgtctattaataagaagatgaacccctcttt 240
|||||

DB 1133 TCATCTCTCTACATCTCGGGAGAATCTCGTATTAATAAAGAGATGAACCCCGTCTTT 1192
|||||

QY 241 agatccaagacaagtgcgactctttcagaaatcagcctgtcacaacatatctaggg 300
|||||

DB 1193 AGATCCAGACAAGTGGGACTTCTTCAGAAATCAGCCTTTGTCAGAGATATCATGAGGG 1252
|||||

QY 301 attagtccaagaaataatttttcaggaaatatagagcctgctcttcgagaagactactc 360
|||||

DB 1253 ATTAGTTCAAGAAAATAATCTTTCAGGAATATAGACCTGCTCTTTCAGAGAGACTACTC 1312
|||||

QY 361 tctttctcggagagactcactatacttttaaaactgaaatcagcttttttgagtaatgg 420
|||||

DB 1313 TCTTTCTTCGAGGAGACTCACTATATCTTTAAACTGAATCAGCTTTTGGAGTAATGG 1372
|||||

QY 421 cgaaccttaacagctgagactttatagaatttggaaacagtagctactcaagaagt 480
|||||

DB 1373 CGACCCTTTAACAGCTGAAGACTTTTAGAATCTTTGGAAACAAGTAGTACTCAAGAAGT 1432
QY 481 ctccaggaatctatgcttttgccttgaaatcccaattaaaaatctacgaagatccaagagg 540
|||||

DB 1433 CTCAGGAATCTATGCTTTTGCCTTGAATCCAATTAATAATGTAGAAAGATCCAAGAGG 1492
|||||

QY 541 acacotctccatagaccattttggagtgcactctccttaataatgaatctacactctgtttac 600
|||||

DB 1493 ACACCTCTCCATAGACCAATTTGGAGTGCACCTCTCTAATGAATCTACATCTGTGTGTAC 1552
|||||

QY 601 cctggaatccccaacotcgcatttttaaaacttttagctcttccagctcttttcccggt 660
|||||

DB 1553 CTTGGAATCCCAACCTCGCAATTTCTTAAACATTTTACCTCTCCAGCTCTTTTCCCGGT 1612
|||||

QY 661 tcataaatctcaagaagacctgcgaatccaaatctcctacactatagcgaagcggagctttcta 720
|||||

DB 1613 TCATAAATCTCAAGAAGACCTCGCAATTCCTAATTCCTAATTCCTAATTCCTAATTCCTA 1672
|||||

QY 721 tcttaaaataatcaaaaaaaacaaatggataaaactctcaaaaaacccctcactactataa 780
|||||

DB 1673 TCCTAAAAAATATCAAAACAAAACAAATGGATTAACCTCTCAAAAACCCCTCACTACTATAA 1732
|||||

QY 781 tcaagtcagggtggaactaaacacgaattacgattcacttccctccogattgcaaacacagc 840
|||||

DB 1733 TCAAAAGTCAGGTGGAACTAAACACGATTACGATTCACTTCACTTCCCGATGCAACACAGC 1792
|||||

QY 841 agcaaaactatttaacacagggaaaaactcaattggcaagggacctccttggggagagcagat 900
|||||

DB 1793 AGCAAAACTATTTAATCAGGGAATACTCAATTCGAAGGACCTCTCTTGGGGAAGACGAT 1852
|||||

QY 901 tctcagaacacctatccaattttacagcttaaggggacactacactcttttgatgtgc 960
|||||

DB 1853 TCCCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACTTACACTCTTTTGATGTGC 1912
|||||

QY 961 aggaacctatgctcactcactcaatatacaataaattccccctcaacaatatgaagcttag 1020
|||||

DB 1913 AGGAACCTATGCTCACCCTCAATATCAATAAATTCCTTCAATATCAATAAATTCCTTCA 1972
|||||

QY 1021 agaagccttagcctagccttagaaggaagctcttctgtctcaactatctcttagccg 1080
|||||

DB 1973 AGAAGCCTTAGCATGAGCTTAGATAAGGAAGCTCTTGTCTCACTATATATCTTTAGCCG 2032
|||||

QY 1081 tgcaaaactgcgactcactcactacataatatttcatagctatccccgaacatcaaaa 1140
|||||

DB 2033 TGCAAAAACCTGCCGATCATCTCTACCTACCAATATTCATAGCTATCCCGAACATCAAAA 2092
|||||

QY 1141 acaagagatggcacaacgccagcttaagcttaaaaaactctttaagaagcgtttagaaga 1200
|||||

DB 2093 ACAGAGATGGCAACACGCCAGCTTAGCTTAAAGAACTCTTTTAAAGAAAGCTTTAGAAGA 2152
|||||

QY 1201 actccaatacactgctaaagatctcgaacatcttaattcttcttctccgcttctcctgc 1260
|||||

DB 2153 ACTCCAAATCACTGTAAAGATCTCGAACATCTTAATCTTATCTTCTCCCGTTTCTGTGTC 2212
|||||

QY 1261 agcaagttcttactagtcacactatacagacagtggaagaaagtttaggggtgc 1320
|||||

DB 2213 AGCAAGTCTTTTACTAGTCCAACTTATACGAGAACAGTGGAAAGAAAGTTTAGGTTCCG 2272
|||||

QY 1321 tatccctattgtcggaaaggaatttgcctcttcccaagcagacctatctcaagggaaatt 1380
|||||

DB 2273 TATCCCTATTGTGCGAAAGGAATTTGCTCTTCTTCAAGCAGACCTATCTTTCAGGGAACT 2332
|||||

QY 1381 ctcttagctacaggaggtgtgttcgcagacttgcctgctctgctatgctgacttcttaagct 1440
|||||

DB 2333 CTCTTTAGCTACAGGAGGATGTTTCGACACTTTTCTGATCTTATGCGATTTCTAACGAT 2392
|||||

QY 1441 ctttgccttcaatcaggttctcctcttatgcaatcaacccatgaagcacttctcctgaat 1500
|||||

DB 2393 CTTTGTCTTATCCATCAGGAGTCTCTCTCTTATGCAATCAACCAATGAAGACTTCTTGAAT 2452
|||||

QY 1501 tctcaaaaaatagaacagagcaagatcccaaaaaacgtctggaattaggtctgcaagc 1560
|||||

DB 2453 TCTACAAACATAGAACAGAGAGAGATCACCACAAAACGCTCGGAATTAGTGTGCAAGC 2512
|||||


```

/genes="CP0577"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38395.1"
/db_xref="GI:7189489"
/translacion="MKRRHSFDSTSKKAVSKAIOKLIKIMETDPSLNVETPNAEI
ESLOIEIKOKLSKOAEDELLEKXCSQFTLSNLEWNASLALSIGSVIELASLK
OLVERIESIGQDQDLOIOLVIEISDKPLSIGTETLSGNLDMNONVIGLLIKENPE
KEMASVYVOTLEPLEKRIGETHAKVATHDVKVSSQFTLSPKMSVAGGRPRGHIDNNG
YRVGLGPKNGEDAVSKDYERYSVQSLTDKVEDKIPKPKGKLLYSQSTSKPLE
GELPLGSLTSGISGTWKSASNGDGFPSALRHKETESDTCFOITISVTLGNQAG
TYTWSLKLVPVFIQIERPEVOLSLVSYEDWLPIDNIFNMSQRPETIPALIGQTM
LAGQYDILELAHQTNQTLMSPCNSRSLQKQTNFENSVPDVFYVHAHSCHWS
GP"
complement(6069. .9497)
/genes="CP0578"
complement(6069. .9497)
/genes="CP0578"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38396.1"
/db_xref="GI:7189490"
/translacion="MKRRSLKILGICLSSIVLGLFIPQLLSTESKYLVSILH
KESGILSCSAEELKISWFGROTARKILGEARDEVFSAERFELDGLLRLIYKKPKG
ITLSWSKINEPASIDHPSVSHLDPSLTYLNDCKIIEHGFITMKVSGSSLSVS
GILSKSEKEMTCVWSEDDQGNIFIESVLPDVSISQAQSVFVAFKIFIASPF
WHLSSYEDIIINSAEATHNDCKISMTASGEGNOIKMLOGHKHKSTFYIVGSSSF
IELKPELASLNCQIIPLSPTISQIHAIVSYAKIPLDITKWKHIEITISQALPEVA
IHPKQNALQLRDRLKLGKTEKESDIYSSVILGASPSHLNGLISIDNKKHLTK
FRLOQALPHYLRALPPOPVINPLOWAYSISNIEGTYRNAHLEADAILNDPLKLG
SCMSGAKNPLFKQGTQHFNKQOILSPHSFYAEARFSGKAQITDNLPPFSG
KITARENELLIHAFQSPNEPIKPTTILIHGFQCSLPSLSNHLAPFLHKLTFS
FHTDGGTQKNGILQENIPDYNPDLNTRILIIPOLLSDSESSTSPSKDLKIQGSG
EFLPSLDITITKYQGVRLSPYFGSGDLNFPVYNPKDQNKLTLLSKFSEALGE
LKLVMDFSMKLSSTQCTLOWESPERSFFKNASCSPICLLHRTANVRDLISLC
PRETKGLSCLTLAAGLGESLEAPLIFYDNVSKETFIINDFKSIRANNIDAKIEY
DLKSCIALPRODSTLAERFSLGQVDHLFSPESREFKQTAWIHIPSPIAGIIPMS
GLIKAGISLAGPRINYSIKNAFRFEGVDINDVSENLOAQIPLILNKSILLRENLT
AHLISNEDVKARLOEFNPLLAGAYSOYPTVIDKONEFLPIRYSPEEPIQOSAT
LDMKISITANTGMWALFQDLTDQKQVSVFTPIFESVQSGSIICKRXDALIDR
IRLAWGKTDIAHDLRFLMTGLIDPEVIRKYPHNTSLATKNFELIKIRGISSEPDWS
SAYARIALKSYISGNPSSSLADKLFLSSLDGSTPPTPVHFFPWKSNFSDIENK"
complement(9494. .10786)
/genes="CP0579"
complement(9494. .10786)
/genes="CP0579"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38397.1"
/db_xref="GI:7189491"
/translacion="MFRKLFPESKKTKGOKORLNNGLLQALIQSIKVLLINEAXKEA
CVLSYGLLTCVPLVFLVFLLSQHLFTNLNKKWELIIPDKPKPIVAIVAAHYATE
SNIGLVGSFFVFCWAGILMLLSLEGLNKIPIFTSWTPISLKRLVSFYVITLVSPMI
FIIVCGSWIITQIMPIQYAKMLLSLESHMTALYISRFVPLLYLLYALFACYAFLPV
AIQTSALISLIGSWIWFQAFPSLOYISFNYFTYGALVALPSPLLYLYTM
YLFQGALFTFIQNRGCTFIIGDKILPSCYLQITSTYILALTTRQNEGLSPITAQF
IAKOSKPIGVSQCLDLEKEGELFPYNNYQPVFNFSELTIKIDIAKLLHREIFKK
FNPLDGLFTFENSQNFQNAKSNENLITSEIRARIK"
complement(10849. .11649)
/genes="CP0580"
complement(10849. .11649)
/genes="CP0580"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"

```

```

/codon_start=1
Query Match          99.9%; Score 1797.4; DB 1; Length 11764;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1798; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acttccccctgctaataactatgctcgagataatgctgcttatgattgcagggtctaggggag 60
Db 1807 ACTTCCCCCTGCTAAACTATGCTTCAGATAATGCTGCTATGATTCGAGGTCTTAGGGGAG 1748
QY 61 aaaaatttcaaaaaactctaglattccggaattcgtatatgcgcaagatatcagtgagg 120
Db 1747 AAAATTTTCAAAAAACTCTAGTATTTCCGGAATTCGTATATGCGCAAGATATACAGTGG 1688
QY 121 aatctgtatccacttctcttcttagcctctccgtagtctccaaagctgcaaggagtcag 180
Db 1687 AATCTGTATCACCATTCTCCTTAGCCTCTCCGTAGTCTCCAGGCTGCAAGGAGTCCAG 1628
QY 181 tcactcctctacatctcggggaactcgcctataataagagatgaaccccgctcttt 240
Db 1627 TCACICCTCTACATCTCGGGGAACTCGCTATTATAATATAAGAGATGAACCCGCTCTTT 1568
QY 241 agatccagacaagtcgacttcttccagaaatcagccttgcataacatatctatgagg 300
Db 1567 AGATCCAAAGACAAGTGGACTCTCTTCAGAAATCAGCCTTGTCAACATATCTATGAGGG 1508
QY 301 attagtccaagaaaataatcttcaggaaatatagagccttcttgcagaagactactc 360
Db 1507 ATTAGTTCAAGAAAATAATCTTTCAGAAAATATAGAGCTGCTCTTTCAGAAAGACTACTC 1448
QY 361 tcttctcgcagcagctcacttacttctttaaactgaatcagccttttggagtaag 420
Db 1447 TCTTTCCTCGGACGAGCTCACTTACTTCTTAACTGAAATCAGCTTTTGTGAGTAATGG 1388
QY 421 cgaccccttaacagctgaagactttatagaaatcttggaacaagtagctactcaagaagt 480
Db 1387 CGAACCCCTTAACAGCTGAAGACTTTTATAGAACTTGGAAACAAGTAGTACTTCAAGAAGT 1328
QY 481 ctcaagaaatctgctttgcttgctgaatccaatataaaatgtacgaaagatccaaagagg 540
Db 1327 CTCAGGAATCTATGCTTTTTCCTTGAATCCAAATTAATAATGTACGAAAGATCCAAAGAGG 1268
QY 541 acactctccatagacacttttggagtcactctcttaataatgactatcacttctgtttac 600
Db 1267 ACACCTCTCCATAGACCACTTTGGAGTGCACCTCCCTTAATGAATCTACACTTTGTGTATC 1208
QY 601 cctggaatcccccaacctgcacttcttcttaaaacttttagctcttccagcttttcccg 660
Db 1207 COTGGAATCCCCAACCTCGCAITTTCTTAAACTTTTAGCTCTTCCAGTCTTTTCCCGGT 1148
QY 661 tcataaatctcaagaacctgcactccaaatctctacattctccgatgcaaacacagc 840
Db 1147 TCATAAATCTCAAGAACCTCGCAATCCAAATCTCTACCTTATAGCAAGCGGAGCTTCTA 1088
QY 721 tcttaaaatatacaaaaaaacaatgataaaactctcaaaaaacctcaactactataa 780
Db 1087 TCCTAAATAATCAACAAAAACAATGGATAAACTCTCAAAAAACCCCTCACTACTATAA 1028
QY 781 tcaagtccagttggaactaaacgatttaagattcacttcaatcccgatgcaaacacagc 840
Db 1027 TCAAGTCAGTGGAAACTTAAACCAATACGATTACGATTCTTCCCGATGCAACACAGC 968
QY 841 agcaaaactatttaacagggaaactcaattggcgaagacacctcttggggagaagcact 900
Db 967 AGCAAAAATAATTAATTCAGGAAAAAACAATTTGCAAGGACCTCTCTGGGAGAACCCAT 908
QY 901 tcctcaagaaacctatccaatttacagtcotaaggggacactaacactcttttgatgctgc 960
Db 907 TCCTCAGAAACCCCTATCCAAATTTACAGTCTAAGGGGCACTTTACACTCTTTTGTATGTCG 848
QY 961 aggaacctcatggctcaccttcaatatcaataaattccccctcaacaatatgaagcttag 1020
Db 847 AGGAACCTCATGGCTCACCTTCANATATCAATAAATTTCCCCCTCAACANATATGAAGCTTAG 788

```

Qy 1021 agaagccttagatcagccttagaagaagccttcttctcaactatattcttagccg 1080
 Db 787 AGAAGCCTTAGCATCAGCCTTAGAATAAGGAAGCTCTGTCTCAACTATATTCTTAGCCG 728
 Qy 1081 tgcataaactgcgatcctccctaccataatcttcaactatctccgaacatacaaa 1140
 Db 727 TGCAAAACTCGGATCATCTCTACCTACCAATATCTATAGTATCCGGAATCAAA 668
 Qy 1141 acaagagatggcacaagccttagcctcaaaaactctttaaagaagccttagaaga 1200
 Db 667 ACAAGAGATGGCACAACGCCAAGCTTACGCTAAAAAACCTTTAAAGAAGCTTTAGAAGA 608
 Qy 1201 actccaaactcactgcataaagctcgaacatcttaattcttctcccttccctcgtc 1260
 Db 607 ACTCCAAATCACCGCTAAGGATCTCGAATCATCTTATCTTTCCTCCGTTTCTCGTC 548
 Qy 1261 agcaagtcttacttagtccaaactctagcagacagtggaagaagcttagggctgc 1320
 Db 547 AGCAAGTCTTTACTAGTCCCACTTATAGAGACAGTGGAAAGAAAGTTAGGGTTCG 488
 Qy 1321 tatccctattgctgaagaagatttgcctctcttcccaagcaactcttcagggaactt 1380
 Db 487 TATCCCTATGTGCGAAAGAAATTTGCTCTCTCCAAAGCAGACCTATCTTCAGGAACTT 428
 Qy 1381 ctcttttagctcagagagatggttcgacagacttgcctcctatgcctatgcattctcaacat 1440
 Db 427 CTCTTTAGCTACAGGAGGATGGTTCGACAGACTTTCCTATCTTATGCGATTTCTAACGAT 368
 Qy 1441 ctttgccttatccatcagagatttccctcttcttgcatacaccataaggaacttctagaat 1500
 Db 367 CTTTGCCTTATCATCAGGAGTTCCTCTTATGCAATCAACCATTAAGGACTTCTAGAAAT 308
 Qy 1501 tctcaaaaactatacaacagacagatcaccataaaacgctcggaattagtgccgaagc 1560
 Db 307 TCTCAAAACATAGACACAGACAGCAATCACCACCAAAAGCTCGGAATAGTGTGCGAAGC 248
 Qy 1561 ttctcttacctagagacatttccatattattgagccgactcaccacagacatttcaatt 1620
 Db 247 TTCTCTTTACCTAGAGACCTTTCATATTATGAGCGGATCTACCAAGCGCATTTCAAT 188
 Qy 1621 tgcctatgaataaaaacttctaatcttagagctctcaccacagaggttggacttcg 1680
 Db 187 TGCTATGAATAAAAACCTTTCTAATCTAGGAGTCTCACCACAGGAGTGTGCGACTTCG 128
 Qy 1681 ttatcctaaggaataatagcactcttcttctcgaactctcgaactgaactgaatt 1740
 Db 127 TTATGCTAAGGAATATTAGCACCTCTTTTATCTCGCAACTTGTCAAGAACTGAATCTT 68
 Qy 1741 atactaaactgggtgccttggcactcttctctctgactgctctctctctctct 1799
 Db 67 ATACTAAACTGGGTGGCTTTGTGGACCTCGTTTCTTCTGACTGCTCTCTCTCTCTA 9

RESULT 4
 LOCUS
 DEFINITION Chlamydothilla pneumoniae J138 genomic DNA, linear BCT 08-DEC-2000 section 1/4.
 ACCESSION AP002545
 AB002545 AB033781 AB033792 AB033793 AB033794 AB033795
 AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
 AB036074 AB036075 AB036076 AB036077 AB036078 AB036079 AB036080
 AB038347 BA000008
 AP002545.2 GI:9956082
 KEYWORDS
 SOURCE Chlamydothilla pneumoniae J138 (strain:J138) DNA.
 ORGANISM Chlamydothilla pneumoniae J138
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothilla.
 REFERENCE
 1 (sites)
 AUTHORS Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.
 TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA

JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
 MEDLINE 20330349
 REFERENCE 2 (bases 1 to 299650)
 AUTHORS Shirai, M.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)
 COMMENT On Aug 31, 2000 this sequence version replaced gi:6172286 gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318 gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162 gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422 gi:8547426 gi:8547430 gi:8978373
 AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)
 AB038345-AB038347: Submitted (14-Feb-2000)
 AB036071-AB036078: Submitted (18-Dec-2000).
 FEATURES
 source
 1..299650
 /organism="Chlamydothilla pneumoniae J138"
 /strain="J138"
 /db_xref="taxon:138677"
 /note="Synonym: Chlamydia pneumoniae (strain J138)"
 complement(1..282)
 /gene="CPJ0001"
 complement(1..282)
 /gene="CPJ0001"
 /codon_start=1
 /transl_table=11
 /product="CT001 hypothetical protein"
 /protein_id="BAA98211.1"
 /db_xref="GI:8978374"
 /translation="MLGKIIIRGLSLIVLICALNVGLIGTHNKLNIIAKLCGGVSTP
 ATQTYIIIGIAGVICLLSFCPCFKSRHSHGSCSSGCGCHSHSDKN"
 573..878
 /gene="gatC"
 573..878
 /gene="gatC"
 /codon_start=1
 /transl_table=11
 /product="Glu-tRNA Gln amidotransferase (C subunit)"
 /protein_id="BAA98212.1"
 /db_xref="GI:8978375"
 /translation="MEQPHLDREIEILLAKASALQSELIQYOTLSAVITSMKCA
 LAIDDDADSCESLPMHVYVNVEDLREDSVTSDFNRBEFLRNVPESLGLVKVPAVK"
 895..2373
 /gene="gata"
 895..2373
 /gene="gata"
 /codon_start=1
 /transl_table=11
 /product="Glu tRNA Gln amidotransferase (A subunit)"
 /protein_id="BAA98213.1"
 /db_xref="GI:8978376"
 /translation="MYRYSALEKAVTLGELTATGVTQHFHRIIEAEGQVAFISL
 CKEQALQELDIKKRSRGEPLKGLAGVPVGIKNDIHVTLGKTTCASRVLENYQPPD
 ATVVERIKKEDGIIILKLMDFAGSTTLYSAPHTNPDLRSVPGGSSGSAAY
 SARFCPVAGSDTGGSIQPAFCGVYGFKPSYGVSRVGLVAFASLDDQIGPLANTV
 EDVALMDVSGRDPKDATSREFFDSFMKSLSTFVPKVIQVPRFLEGLRDDIRENF
 FSSLAIFEGEGHLDVDELILSHAVSIYIILASAAATNLFARFDGVRYGYSRQAHT
 ISQLDLSRGEFGKEVMRILLGNLYLSAERONYTKATRAKIVKAFRTAFKFC
 EILAMPVCSAPAEIGEILDEVTLYLQDIYTVAMNLAYLPAIAVPESFSGLEPLGLQ
 IIGQGGQDQGVQGVSYFQEAQIKQLFSKRYAKSVLGGQS"
 2370..3836
 /gene="gatB"
 2370..3836
 /gene="gatB"
 /codon_start=1
 /transl_table=11
 /product="(pet112) Glu tRNA Gln amidotransferase (B subunit)"
 /protein_id="BAA98214.1"

```
/db_xref="GI:8978377"
/translation="MSAAVADWESVIGLEVHELVETLNTASKLFSSALNRFCDPENTNIST
VCTGLPGSLVNLGSAVEKAVLPGCAVEGEISLRSRDRKSYFIPDSNPFOITOFEPH
PLIIRGRRIKAVOGEERYFELAQTHIEDAGMLAHFGEFAGVDYNRAGVPLLEIVSKP
CMFCPEDAVATATSLSLDDYIGISDCNMEGSIKFDVNVSRPGKSPCLRNKVKIKN
MNSFAQAALAEAKQKQIDVYLNQPNKDEKLVIPAAATYRWDPKKTIVLMRLKESAE
DYKPEPDLPTQLTQTESYIERIKTLPELVDYKHYIOEYLSBDIASIILISPKNI
ATFFEVACKDCNFRSLSNWVSVEFGCKTLGVKLPSGSIFFEGVAQLVNAIDGVI
TKGIKAKETADIMMESCKNPEILKEKPELLPMSEDEGELQKIIAEVVLNAPESIVDYK
NGKTKAGFLVGQIMKRTAGKAPPRVNNELLLELDGK"
4127..6895
/gene="pmp_1"
/gene="pmp_1"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family"
/protein_id="BAA98215.1"
/db_xref="GI:8978378"
/translation="MRFSLCGPLVFSEFTLLSVFDTLSLSTISLTPDSPHGDSQNA
ERSYVAGQGVYSLTGDVSTISNVDNALNKACFNVTSGSVTFAGNHGILFNNISSGT
TKGAVLCCODPOATARESGFTSLFIQSPGDIEQGLYSKKNALMLNNYVYRFEQN
QSKTKGGAISGANVTIVGNDSYFQNAATFGGAHSSGPIQIAVNAQAEIRFQNTA
KNGSGALYSDGIDIDQWYVLFRENEALTTATGKGAVCCCLPTSGSTPVPVITFS
DNKQLVFERNHSIMGGGAIYARKLSISSGPTLFINNISYANSQNLGAIAIDTGEI
SLSAEKTITFOGRTSLFPLNGIHLLQNAKFLKQARNYSIEFFDPTISAGDGTQ
LNIINGDKNEYTGTILFSGEKSANDPRDFKSTIPONVNLGAGLYIKEGAETVSK
FTQSGSHVLNIDGTGLKIASKEIDAITGLAIDTIDSSSTAAVTIKANTAKQISVID
SIELISPTGNAYEDLRMSOTFELLSPGAGSVIVTAGDPLVPSHYFGQGNKL
AWTGTGNKVEFEWDKINKPRPEKGNLVPNLWNAVDVRSLSMOVQETIASLIQTD
RGLMDGFHFHVSASEDNIIRHNSGGYVLYSYNNEITPKHYTSMATFSQLSFKDQY
AVSNNEYVGLSYLYQYTTISIGNIFRYASNPVNVGILSRFLQNLMTIPIFLCAY
GHATNDKMTIYANFPMKSNWNCWATECGSGMPLLVPENGRLFOGALPMKQLVY
AYQDKEFTTAGDRFRSNGSLTSISVPLIGRFKIALSODVLYDPSFYIPIQIRKP
SCEAALVYSGDSMLVPAAHVSRHAPVSGTGRVHFNDYTELLORSIECRPHARNYI
NCGSKKFRF"
complement(7138..7293)
/gene="CPj0006"
complement(7138..7293)
/gene="CPj0006"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98216.1"
/db_xref="GI:8978379"
/translation="MQEPLRSALLERLEWLVLLGVSPETTRSTPEKDNQLPKDSR
NRTLESL"
7515..10169
/gene="CPj0007"
7515..10169
/gene="CPj0007"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98217.2"
/db_xref="GI:9956083"
/translation="MFSELVPIPLDTTSSLSLTDDEGNPQSMRKLRIILAIIVLIAL
SILILAGVLLVLPVLPGLSVISPPAGMACALGCVMLALGIDVLLKKRVPVILAS
VTTPVTSPRSGISIGADSTIRSLPTYLDEHPQSMRKLRIILAIIVLIVFSILILIA
SGVLLTVAIPGLSIISSPAEMGACALGCVMLALGIDVLLKKRVPVLPVAPRPEV
VIDIDEEISIRLQOEAEALARLPPEMSAFEGYIKVVESHLENKMSIPLYOGLIEBK
KHQIVRSRLKAMPPELDJRIREFEEFFLSAKRLIDLDTLVERKILTEQLER
NNLRKAFSLYQDSIFKKIINDFEKAWPMILSKISICRFTIIFENHEGVAKSLHK
NAVLLERVIYLSQVSDIGMSAKMKILHGNPFFSLEDNKTIMKHAEMLESLS
YKRVFALSDEVNVTSPDPKWDLSGIPCDRALSEISRDQWOKKHALKHQESLYTO
ARDLDTOSKENQEKLEAOEYIISWVRVKFETEVQERLQALQKLYPNILREE
ETTQGTQVTPVQGTASSDITDILGRIEVSSREDNQESCIVKLVLSHEVMSWVK
QEVGPKKKEFQDMSIERFTEHIELEVLQKDYSKHLSFKVKNKKEVQAKFRL
KVLSDLEGILQTESLSLTOELPILATRALEKAVFAGSLCCALASKAPFRL
DPRFQSDTQRAUTLRLOAKASLEEEKRFSNLENDTAEERLLKESKOTFERAGL
GVLYREAVESYDLSLNTWEGTPESEKVIYSWYLNLYNEEKRRKRLVEMTQYR
DFKMALEAMQFNEALLQEELISQAPSE"
10645..11358
/gene
```

```
CDS
/gene="CPj0008"
10645..11358
/gene="CPj0008"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98218.1"
/db_xref="GI:8978381"
/translation="MISGLLLVRRREVTVRSEIIPRGVSVTSEEPALAEAKOKEPE
TKKILDRLPKELDQDLYIOEVFAELRLKDPKYEDRGLLTKAEKILRVFDVVKDDM
SEFLDIORVLEEAAYVEHCQDPLENTAYEIESSOELDYACGVCGYLPSGDARADR
LAKRSKEYMDRPMRTWKSWEASVMDHSYGVALELFKKAAGVLEESVYKILFKSTRD
AFTECEKAKIQDRGRFKWL"
11485..12792
/gene="CPj0009"
11485..12792
/gene="CPj0009"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98219.1"
/db_xref="GI:8978382"
/translation="MWLDRYADKFIILREKEEKMERHELFRATWVRKASGHAKAKAA
EKERSNENOKVADVEKWLKSLGAEFRNQESRRARERLQLTILPEVSVEVLER
QRTKYNLENLIADIEKKYHHCVRQEHWKEVNEKAEYRENGEKVLSAEVSECLQ
RLEDCLETWSKKLTAAEESVFEMKFDATKGNKVLSDVNRLEILCEADAEMIFRIE
EMILTRWELPLFLFMKNTPEKASLOYNCKEMAKHVEPOCKESPITVRSQERLERN
QDQATYVNCQERLQGSLEKSVRTRCDRLREQMKHFEVQGLNFINEELLWVGAELF
TQARDLVATVPYMEFYLOYHNIKREKVRSONMAKTEREYRIQATFGQVWKEDLLAED
TILKDEEDVLLDRDMLLDRERKNRRLCNKIAAAQQRVKGF"
13104..15428
/gene="CPj0010"
13104..15428
/gene="CPj0010"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
Query Match 99.9%; Score 1797.4; DB 1; Length 299650;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1798; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 actccccccctgtaaacatgctcagaataatgctatgattgacagtgctagggag 60
Db 233705 ACTTCCCCCTCGTAAATATGCTCAGATAATGCTGCTATGATTGTCAGGTCTAGGGAG 233764
QY 61 aaaaatttcaaaaaacttagtattccggaaattcgatatgcgaagatatcagtggg 120
Db 233765 AAAATTTCAAAAAAACTTAGTATTCGGAAATTCGTATATGCCAAGATATCAGTGGG 233824
QY 121 aatctgtatcacattctccttagctctcctgtagtccctccaaagctgcaagagtcag 180
Db 233825 AATCTGTATCACATTCTCTTAGCCTCTCCGTAGTCCCTCCAGGCTGCAAGGAGTCCAG 233884
QY 181 tcactcctctacatctcggggagaaactcgctattataataagaagataaacccgtcttt 240
Db 233885 TCACCTCTCTACTATCGGGAGAACTCGCTATTAATATAGAGATGAACCCGTCTTT 233944
QY 241 agatcaagaagaagtcgactcttttcagaataatcagcctttgcaaacatctatcaggg 300
Db 233945 AGATCAAGAACAAGTCGCACTCTTTTCAGAAATCAGCCTTTGTCAACATATCTATCAGGG 234004
QY 301 attagttcaagaataatctttcaggaataatagagcctctcttcagaagactactc 360
Db 234005 ATTAGTTCAAGAAATAATCTTTTCAGAAATATAGAGCTCTCTTCGCAAGACTACTC 234064
QY 361 tcttctcgcagagactcaactatacttcttaaaactgaaatcagctcttttcaggaagtaag 420
Db 234065 TCCTTCTCGGAGGACTCACTATFACCTTTTAACTGAAATCAGCTTTTGGAGTAATGG 234124
QY 421 cgacccttaacagctgaagactttatagaattcttggaaacaagtagctactcaagaag 480
Db 234125 CGACCCCTTAACAGCTGAAGACTTTTATAGAATCTTGGAAACAAGTACTACTCAAGACT 234184
```

QY 481 ctccaggaatctatgcttttgccttgatcccaattaaatgtacgaagatcccaagagg 540
 Db 234185 CTCAGGAATCTATGCTTTTSCCTTGAATCCAAATTAATAATGTACGAAGATCCAGAGGG 234244
 QY 541 acactctccatagacacatttggagtgactctcctaatgaatctaaacttggttac 600
 Db 234245 ACACCTCTCCATAGACCAATTTGGAGTGCACCTCTCCCTAATGAATCTACACTGTGTGTAC 234304
 QY 601 cctggaatccccaacctcgcatcttttaaaacttttagctcttccagttcttttcccg 660
 Db 234305 CCGTGAATCCCCAACCTCGCATTTCTTAATAACTTTTAGCTCTCCAGTCTTTTCCCGGT 234364
 QY 661 tcataatctcaagaacctgcaatcccaatctctacotatagcaagcgaggtttcta 720
 Db 234365 TCATAATCTCAAGAACCCTGCAATCCAAATCTCTACCTATAGCAAGCGAGGTTCTTA 234424
 QY 721 tcttaaaatatacaaaaaaacaatggataaaactctcaaaaaacccctactactataa 780
 Db 234425 TCCTAAAAATATCAAAACAAAAAATGGATAAAACTCTCAAAAAACCTCTACTATAA 234484
 QY 781 tcaaaagtcagggtggaactaaacgattagattcacttctcccgatgcaaacacagc 840
 Db 234485 TCAAAAGTCAGGTGGAACCTTAAACGATTAGCATTCATCTCCGATGCAAAACACAGC 234544
 QY 841 agcaaaactatttaatacagggaactcaattgcaaggaacctcttggggagagcgat 900
 Db 234545 AGCAAAACTATTATCAGGGAACCTCAATTGGCAAGGACCTCTTGGGGAGAACGCAT 234604
 QY 901 tctcagaacacctatccaatttacagtcgaaggcgacttaactctttttagtgcgc 960
 Db 234605 TCCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACTTTACACTCTTTTGTATGCC 234664
 QY 961 aggaacctcaatgctcaacttcaataataaattccccctcaacaatatgaagcttag 1020
 Db 234665 AGGAACCTCATGGCTCACTCTCAATATCAATAAATTCOCCTCAACAATATGAAGCTTAG 234724
 QY 1021 agaagccttagcatcagccttagaagaagcctctgtctcaactatattttaggcgc 1080
 Db 234725 AGAAGCCTTAGCATCAGCCTTAGATAGGAAGCTCTTGCTCAACTATATTTCTTAGCGCG 234784
 QY 1081 tgcataaactgcgcatctctcactcaataatattcattagctatcccaacatacaaa 1140
 Db 234785 TGCATAAATCTGCGATCATCTCTCTACCTACCAATAATTCATAGCTATCCGCAACATCAAA 234844
 QY 1141 acaagagtgccacaaagcgaacttacgtcaaaaaactctttaaagaagcttttagaaga 1200
 Db 234845 ACAAGAGATGGCAACGCCAAGCTTACGCTAAAAAATCTTTAAAGAAGCTTTAGAAGA 234904
 QY 1201 actccaaatcactgctaaagatctcgaaactcttaactcttattcttcccgctttccctgc 1260
 Db 234905 ACTCCAAATCAGCGCTAAAGATCTCGAATCTTAACTCTTATCTTTTCCCGTTTCTCTGTC 234964
 QY 1261 agcaagttcttactagtcacactttacagaaacagtggaagaaagttaggttcgc 1320
 Db 234965 AGCAAGTTCTTTTACTAGTCCAACTTATAGGAAACAGTGGAAAGAAATTTAGGGTTCG 235024
 QY 1321 tatccctatfctggaaagaaatttgcctctctccaagcagacctatcttcaggaactt 1380
 Db 235025 TATCCCTATTCTGGAAAGGAATTTGCTCTCTCCCAAGCAGACCTATCTTCAGGAACTT 235084
 QY 1381 ctcttagctacaggaggtggttcgcagactttgtgactcttagcttagcttagcttagct 1440
 Db 235085 CTCTTAGCTACAGAGGATGGTTCGACACTTTGCTGATCTTATGGCAATTTCTTAACAT 235144
 QY 1441 ctttgccttatccatcaggagttcctctttagcaatcaaccataaaggaacttcttagaag 1500
 Db 235145 CTTTGCCTTATCCATCAGGAGTTCCTCTTATGCAATCAACCAATAGGACTTCTTGAAT 235204
 QY 1501 tclacaaacatacaacaagagcagatcaccaaaaaacgtcgtgaattagtctcgaagc 1560
 Db 235205 TCTACAAACATAGAACAGAGCAAGATCACCAAAAAACGCTCGGAATTAGTGTCCACAGC 235264

QY 1561 ttctctttacctagagacctttcatattattagccgatctaccacacacatttcaatt 1620
 Db 235265 TTCTCTTTACCTAGAGACCTTTCAATATTATTAGCCGATCTACCACACCACTTTCAATT 235324
 QY 1621 tgcctatgaataaaaaacttttctaacttagaggtctcccaacagaggttggactccg 1680
 Db 235325 TGCTATGAATAAAAAACTTTCTAATCTAGAGTCTCCACCAACAGAGTGTGGACTCCG 235384
 QY 1681 ttatctcaagaaataatagacacctttttaaactctcgaacctgtcagaactgaattt 1740
 Db 235385 TTATCTAAGGAAATATAGCACCTCTTTTAATCTCGAAACTGTCAAGAACTGAATCTT 235444
 QY 1741 atactaaactgggtgccttttgggcaacctgttcccttcgactgtctctctcteta 1799
 Db 235445 ATACTAACTGGGTGCCTTTGTGGCACTCGTTTCTTCTGACTGCTCTTCTCTCTA 235503

RESULT 5
 AX349501
 LOCUS AX349501 1599 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 24 from Patent W00202606.
 ACCESSION AX349501
 VERSION AX349501.1 GI:18615357
 KEYWORDS
 SOURCE Chlamydomophila pneumoniae.
 ORGANISM Chlamydomophila pneumoniae
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE 1 (sites)
 AUTHORS Ratti, G. and Grandi, G.
 TITLE Immunisation against Chlamydia pneumoniae
 JOURNAL Patent: WO 0202606-A 24 10-JAN-2002;
 Chiron S.p.A. (IT)

FEATURES
 source
 1. 1599
 /organism="Chlamydomophila pneumoniae"
 /db_xref="taxon:83558"

BASE COUNT 509 a 389 c 261 g 440 t
 ORIGIN

Query Match 88.9%; Score 1599; DB 6; Length 1599;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 atgcgcaagatatacgtgggaactctgtatcaccaattctcttagcctctccgtagctc 160
 Db 1 ATGCGCAAGATATCAGTGGAAATCTGTATCACCATTCTCTTAGCTCTCGGTAGTCTC 60

QY 161 caagcgtgcaaggagtcagtgactctctacatctcggggagaaactcgctattataata 220
 Db 61 CAAGCCTGCAAGGAGTCCAGTCACTCTTACATCTCGGGGAGAACTCGCTATTAAATA 120

QY 221 agagatgaaccccgcttcttagatcccaagcaagtcgactcttccagaactcagcctt 280
 Db 121 AGAGATGAACCCCGCTTCTTAGATCCCAAGCAAGTGCAGTCTCTTTCAGAAATCAGCCT 180

QY 281 gtcaaacatctataggggattagttcaagaaaaataatcttcaggaataatagacct 340
 Db 181 GTCAACATATCTATGAGGGATAGTTCAGAAATAATCTTTCAGGAAATATAGACCT 240

QY 341 gctctgcaagaagactactcttctccgcgacgactcaactataacttttaaacitgaa 400
 Db 241 GCTCTTGCAGAAGACTACTCTTCTTCTCGAGGAGCTCACTTATCTTTTAAACATGA 300

QY 401 tcagcttttgagtaatggcgaccccttaacagctgaagactttatagaactttggaat 460
 Db 301 TCAGCTTTTGGAGTAATGGCGACCCCTTAACAGCTGAAGACTTTATAGAACTCTGAAA 360

QY 461 caagtagctactcaagaagctcaggaactatgtcttttgcctgaatccaataaaaa 520
 Db 361 CAAGTAGCTACTCAAGAAGTCTCAGGAATCTATGCTTTTGGCTTGAATCCAATTA 420

QY 521 glacgaagatccaaggggacacctctccatagaccattttggagtgactccta 580

Db 421 GTACGAAGATCCAAAGAGGACACCTCTCCATAGACCACTTTTGGAGTGCCTCTCTAAT 480
Qy 581 gaattacactgtgtgtacccgtggaatcccaactcgcatttttaaaacttttagct 640
Db 481 GAATCTACACTGTGTGTACCTGGAAATCCCACTCGCATTTCTTAAACATTTTAGCT 540
Qy 641 ctccagctttttcccgcttcaataatcgaagaacccctgcaatcccaactctcact 700
Db 541 CTTCAGTCTTTTCCCGCTTCAATAATCTCAAGAACCTCGCAATCCAAATCTCTACT 600
Qy 701 atagcaacgagcttct 760
Db 601 ATAGCAACGGAGCTTTCTATCTCTAAATAATCAACAAACAAACATGATATAAACTCTCA 660
Qy 761 aaaaacccctactactataatacaagtcaggtggaaactaaaacgattagcattcact 820
Db 661 AAAAACCTCTACTACTATATCAAGTCAAGTGGGAAACTATAACGATTACGATTCACTTC 720
Qy 821 attccgatgcaaacacagcagcaaaactatttaacagggaaaactcaattggcaagga 880
Db 721 ATTCCGATGCAACACACAGCAGCAAAACTATTTAATCAGGAAACTCAATTTGCAAGGA 780
Qy 881 cctctctgggagacgcttctcctcaagaacccctatcccaatttcaagcttcaagggcgc 940
Db 781 CTTCTTTGGGAGACGCTTCTCTCAAGAACCTTATCCAAATTTACAGTCTAAGGGCAGC 840
Qy 941 ttacactctttgatgtcgcaggaacccctcctcctcctcctcctcctcctcctcctc 1000
Db 841 TTACACTCTTTTGTATGTCGAGGACCTTCATGGCTCCTACCTCAATATCAATAATTCCT 900
Qy 1001 ctcaacatagacttagaagaaccccttagcatcagccttagatagaagagctctgtc 1060
Db 901 CTCACCAATATGAGCTTTAGAGAGCCCTTAGCATCAGCCTTAGATAAGGAAAGCTCTGTC 960
Qy 1061 tcaactatatcttagcgtgcaaaaactgcgcatctctcctcctcctcctcctcctcct 1120
Db 961 TCAACTATATCTTAGGCGTGCAGGAACTGCCGATCATCTCTACCTACCAATATTCAT 1020
Qy 1121 agctatccggaacatacaaaaacagagatggcaaacgccaagcttaagcttaaaactc 1180
Db 1021 AGCTATCCGAGACATCAAAAACAGAGATGCGACACGCAAGCTTACGCTAAAAAATCT 1080
Qy 1181 tttaagaagctttagaagaactcctcaactcactgctaaagactcgaacacttaactct 1240
Db 1081 TTTAAAGACCTTTAGAGAACTCCAATCACTGCTAAAGATCTCGAACATCTTAATCTT 1140
Qy 1241 atcttcccggttctcgtcagcaagttcttactagtcctcctcctcctcctcctcctc 1300
Db 1141 ATCTTTCCCGTTTCTCTCGTCAAGTCTTTACTAGTCCAACTTATACGAGACAGTGG 1200
Qy 1301 aaagaagtttaggttgcgtatccctctgctcgaagaagattgctcttctcctcctc 1360
Db 1201 AAGAAGTTTAGGTTGCTATCCCTATTTGTCGGAAGAAATTTGCTCTCTCCAAAGCA 1260
Qy 1361 gaactatctcaggaactctctttagctcagagagatggttcgagactttgctgat 1420
Db 1261 GAACCTATCTTCAGGAACCTCTCTTTAGCTACAGGAGGAGTGTTCGAGACACTTTGCTGAT 1320
Qy 1421 cctatggacttctaacgatcttcttctcctcctcctcctcctcctcctcctcctcct 1480
Db 1321 CCTATGGCATTTCTAACGATCTTTGCTTTATCCATCAGGAGTCTCTCTTATGCAATCAAC 1380
Qy 1481 cataagacttctcagaactctcaaaaacatagacaagcagatcccaaaaacgc 1540
Db 1381 CATAGGACTTCTAGAAATTTACAAAACATAGAACAGAGCAAGATCACCAAAACGC 1440
Qy 1541 tcggaattaggtcgcagcttctcttctcctcctcctcctcctcctcctcctcctcctc 1600
Db 1441 TCGGATTAGTGTGCGAAGCTTCTCTTTACCTAGAGACCTTTCATATTATTAGCCGATC 1500
Qy 1601 taccagcagcatttcaattgctatgaataaaaacttctcctcctcctcctcctcctcct 1660

Db 1501 TACCACGACGCAATTCAATTTCCTATGAATAAAAAAACTTCTTAATCTAGGAGTCTCACCA 1560
Qy 1661 acaggagtgttgacttccgttatgtctaaagaaaaattag 1699
Db 1561 ACAGGAGTTGTGGACTTCGGTTATGTCTAAGGAAATAG 1599
RESULT 6
AE001293 11944 bp DNA linear BCT 30-OCT-2000
LOCUS Chlamydia trachomatis section 20 of 87 of the complete genome.
DEFINITION AE001293 AE001273
ACCESSION AE001293.1 GI:3328597
VERSION AE001293.1
KEYWORDS Chlamydia trachomatis.
SOURCE Chlamydia trachomatis.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 11944)
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
TITLE Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
JOURNAL Science 282 (5389), 754-759 (1998)
MEDLINE 9900809
PUBMED 9784136
REFERENCE 2 (bases 1 to 11944)
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 9920606
PUBMED 10192388
REFERENCE 3 (bases 1 to 11944)
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
source
1..11944
/organism="Chlamydia trachomatis"
/strain="D/UW-3/CX"
/db_xref="taxon:813"
complement(105..878)
/gene="CT192"
complement(105..878)
/gene="CT192"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67784.1"
/db_xref="GI:3328598"
/translation="MFPSGVSVELPSTCLISKTRKEFVRVSGQEAQNTLSWTIR
PRIPISIOGSEVSLALVLTAVGACAAAGVAFVCLVILGVLGVLATGILL
AVERPCHLRSSMEKVIITKQLEFPIQEQATPLIEASTCEPGIPLSGPEVQ
OBLFVILKQDLDLSHAPKXIYGVSHVVLKAGIKGRNGERLLEGGIDTDQNFVRVD
ELFLLUGKENVRLDGCCKVLPLPKSKSESINNLVSND"
/gene="tgt"
complement(905..2023)
/note="CT193"
complement(905..2023)
/gene="tgt"
/codon_start=1
/transl_table=11
/product="Queuine tRNA Ribosyl Transferase"
/protein_id="AAC67785.1"
/db_xref="GI:3328599"
/translation="NALREFEILHQSKSRARVRIETAHGYIDTFAFVPVATNGALG
VLDHNSIPLMFCNTYHLIVHPGAEIAAAGGLHOFIGNAPITDSCGFOIFSAYGS
VAEIKSCGKKGGNTIIKVDNDGVHFKSYDGRKLFSLSPESVQAQKDLGAILPL
DELLPFHADPTYPHQSSQFTYWEKRSIDHLKNPFIQSVGIYHGGTFPQRLKGLK

```

FVEDLPFGSAGISGLKQIDIVEVGVTAANLSAERPHLLIGDIPSIWATVGF
IDFSOSSYTKAARGMILTSQGLKINNRYSSDLNP:IEPGSCSLACSGQITRAYLR
HLFKVHEPNAGIWAS:IHNMHHMKQYMR:REGILINDRI"
complement(2150..3562)
/gene="mgte"
/note="CT194"
complement(2150..3562)
/gene="mgte"
/codon_start=1
/transl_table=11
/product="Mg++ Transporter (CBS Domain)"
/protein_id="AAC67786.1"
/db_xref="GI:3328600"
/translation="MOSKTSHLDELCEKLEAFDTLTAGEHSQDLTSIVSYNPIDL
AYAVSCPDSRSITLYKNLSDIASKIAPILINTDSASPAWAFRNLSGECALIEQMP
DEAIWLDLDPDRYRILDLIDVKKALKIKDLOKHGRNAGRLMNEFFALMETTV
KEVAICINNPGIDLRVLVDFKGLQGVTDKRSIIASPEMLKQIMRPIEHVL
AUTREEDVLYVERKVALPVVDNEENELIGAITYEDVETIEDIADETIARMATGE
DVGHDCHVQRELLRAPMLLITLCAGVISASVMAYFOKIAPTLAMWIFRPLVNGL
SONGVQCSITILVRSMTGTLSGRRRETILKEMSIGLTCGVALGILGLVYCCMGCL
GIGLFATGGVQLGVTVSGILGASLTATTLGVLSPPFFAKIGVDPALASGP:IVTALND
IMSVVILLITGLNVLFRK"
complement(4001..5092)
/gene="CT195"
complement(4001..5092)
/gene="CT195"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67787.1"
/db_xref="GI:3328601"
/translation="MVSMSLNLPPAEVRLPVTAKSPFNSSAWRATQKRIRGCSLNAS
SYRSPSARVISEFVGVVLVGLGICFERIAYLANSEVITTFMFAIAILLFFLGGGLAI
LFRITAGKVDTGCKI:TOPEISQWERNILCEKEGIRPIQEPDMVDSLLDRGTSG
IAPVYTPNDATVVICVYSILLFIFSIWMLYNIRFFIYFVILFQVRYQNTD
IPKERVVCSDIVRENTSLQAVKAPFYGAVCYLANLYLLNPLSGRWVLASLERDW
NDVIRSGVWGIFCEKNYMEGGTSGLGQNAWYLLGCFQFQVQLFKDKGVIIISA
RPSIQSPSEKEYLASFLYGAVPGRLAGF"
5316..5636
/gene="CT196"
5316..5636
/gene="CT196"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67788.1"
/db_xref="GI:3328602"
/translation="WATTVNENYSLPECEKMVSSAPLAAHSFFNHIKLLIPLLVGVFC
ILGLAILTGVTITPISAAASYFLSLGVTLIVGAGLCAAFKRPLFSVTQSKASTLLH
ISCQ"
5712..6728
/gene="gcp_1"
/note="CT197"
5712..6728
/gene="gcp_1"
/codon_start=1
/transl_table=11
/product="O-Sialoglycoprotein Endopeptidase"
/protein_id="AAC67789.1"
/db_xref="GI:3328603"
/translation="MLTILGESSCDETSVLQNGKILANKIASQDIHASYGQVPEL
ASRAHLQTFPELLTAATQASGVLEDIELISVANTPGILGALSIGVNFAGLASGLKR
PLTGVNVAHLVIAACMEAPATOPPALGIAISGHTSLFLMPDATTPELLIGKTRDDAI
GETFDKVAFLCLPYPGCKLELEAREGADAFAPARVSGYDFSGGLKTVIYAL
KGNNSAKAPPEVSEYQKNRIASQKAVFTIAQKLPDIDVITKTFSCESILVGGVYAN
NSVFERLLNQ:ICSLPIITFPSSQLCSDNAAMIAGLGERLFCNRTHVSKEVIPCARYQE
SACS"
6692..8248
/gene="oppA_3"
/note="CT198"
6692..8248
/gene="oppA_3"
/codon_start=1

```

```

/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAC67790.1"
/db_xref="GI:3328604"
/translation="MRKISVGICLLLALATSGCSSSNATHRSPATHVAVSVKDDP
RTFDPREVRLSLDINLIHHLYEGLVQTPSPGEVFPALAESFSLSEDKKTYTNLKKAF
WSNGDLITAHDFVSNVDLQNRVASIYSFAFLP:IDVKNOSGFFAKDDHTLVINLLTP
TPHFLKLLTPVFPVSHQHOIRKEEKSLPISTGAFFLKEKKDRMLKLEKSPYYNK
DOVAQVEIGIHIIPDOQTASALNQGLDWQGLPWGHSIPQETLATINKRARPSDI
SGTSLTFTNAKPPESHKLRQALSIVLNKEALASIAFKPAKHLPAHLHTYPEQS
YKQOEAITAKSLEELAEULNMTIEDLEKYPITFSATSTNMSIAQMLRDQWRSLG
ITPFCGKYALLQDNLGNFTFMSIGMFAFSDPLAFLSIF:SSKGVKPYALQDPQF
DQILISIEYKKNPQRSALISEASLYTERQNVIEPLYHDVFHYTTNNKLSFVLHPSG
LVDMYAKNS"
8407..9348
/gene="oppB_1"
/note="CT199"
8407..9348
/gene="oppB_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67791.1"
/db_xref="GI:3328605"
/translation="MLSYIKRRLLFNLLSLWVVTLPFFIITKIPGDPENDENGNTLS
SETLALKNRYGLDKPLFTQYLIYKCLLTDGESLIYKDRIVISIIAAALPSSAIL
GLESICLSLFGSITLGIILAAFYKKSGRITFFSSVQISVPAFVGAFOYVFAIKS
CLPIACWGNFSHILLPSIALAITPMATITQLTICASVSAUKKDYVLLAYAKGLSPFKV
LKHILIPALFPVSIYSAFLITMTITFTSIENTLFCIPGLGKWFICISQKROYPTITG
LSVFYGAFFMTLSICDQLQAWIDPQIRYSYKERSK"
9380..10225
/gene="oppC_1"
/note="CT200"
9380..10225
/gene="oppC_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67792.1"
/db_xref="GI:3328606"
/translation="MFKSSSTHWRYIRTNKMLVGLCLTILAVILVLAATFLPYLDYE
RTFPEHALQSPSKAFPGDNNLGRCLMCLATQGLIRLSLAVSATTIDVCLGLWSTL
ALASKGVADIMGRITEILSIPRIPIVILLIVFNHGLPILILAMTITGWIPIARI
YQGLLEKNKEFVLSARALQASTFHLKRKHLPLNSLIGPILSTLITIPNAVYTERFIS
FLGIGIQQPVASLGLTVKEGHSILAYHPWLFIPSFNFIIVSVFNCIGEGRLTKLE
ENTLV"
10218..11051
/gene="oppD"
/note="CT201"
10218..11051
/gene="oppD"
/codon_start=1
/transl_table=11
/product="Oligopeptide Transport ATPase"
/protein_id="AAC67793.1"
/db_xref="GI:3328607"
/translation="MSEDLKIDNLVSVKDSNRQLVNHLSTIKRQSMALVGENGS
GKTTVSKAVGLFPDNCYTQSGRILYSTDIRLSRQRTIRGKKIATIPQNAMGTL
TPSMRYCAQVETLHHEHDMKSEAFSKARELLESVHIESPORCLQLYPFELSGMQ
RVSTALANVELI:IADEPSTALDSISQAQVLRVLTQIHNHSTALLITHNLALVS
ELCEEMAIRYEI:VEQGPVQELLHSPSHEPTQQLIRAIKPIPSPYLSKPEPLATTA
Y"
11066..11809

```

```

Query Match      24.9%; Score 448; DB 1; Length 11944;
Best Local Similarity 56.7%; Pred. No. le-98;
Matches 976; Conservative 0; Mismatches 695; Indels 51; Gaps 6;

QY 1 actccccccgtgtaactatgctcagataatgctgctatgattgctgagggag 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6589 ACTTCCTCTCTCCAGTATGTTTCAGACACGCTGCAATGATAGCGGGTTAGGAGA 6648
QY 61 aaattttcaaaaaactctatttccc---ggaaatttcgtatatgcgaagatcagt 117

```


Db 6649 GACTGTTTTCATCGACACATGTTTCTAAGGAGGTTCATTCATCGCGCAAGATATCAGT 6708
 QY 118 ggaatctgtatcaacattctcttagcctctcgttagctctcgaagctcgaaggagtc 177
 Db 6709 GGGAACTGCTGCTCCTAGCATATGACAACTCTTGGATGTTCAAAATCTCTCTTAACGC 6768
 QY 178 cagtcactctctacatctcgggagaaactcgtattataagagatgaaccccgcttc 237
 Db 6769 AACCCATCGGCTCCAGCTACTCACAGATGCTGTAAACGTAAGAGATGATCTCCGAC 6828
 QY 238 tttagatcaagaagaagtgcaactctcttcagaatcagccttgtaacacatctatga 297
 Db 6829 ATTGTGATCTCGAGAGGTTCCGCTCTCTTCTGATATCAATTTGATTCATCATCTCTATGA 6888
 QY 298 gggattagttcgaagaaataatcttcaggaaatataagcctcgtctcttcgagaagacta 357
 Db 6889 AGGATTGGTACAGAA--ACTCCTCTCGAGAGTCTTCCCTGCTTTAGCGAGAGTTT 6945
 QY 358 ctctctctctcggaggaactcacttatcttttaactgaatcagctttttggagtaa 417
 Db 6946 CTTCTTATCCGAGAGATAAAAAAAGTTATCTTCAACTTGAATAAAGCTTTTGGAGCAA 7005
 QY 418 tggcgaccttaacagctgaagactttatagaactcttgaaacaagtagtactcaaga 477
 Db 7006 TGGAGATCTTATACCGTCTATGATTTGTCTCTCGAATGATGTGTTTACAAATCG 7065
 QY 478 agtctcaggaactctatgctttgcttgaaatccaatataaatagtacgaagatccaaga 537
 Db 7066 TGTCGCTAGTATTTATTTTCGCTTTTCCTCTATTGA----- 7104
 QY 538 gggacacctctcatalagacaatttggagtgcaactctcactaatgaatcacaactttgt 597
 Db 7105 -----CGTGAATAAGGATTCGTGATTTTTCGCAAGATGATCATACTCTTGTAT 7155
 QY 598 taccttggaatcccaactcgaactctttaaacttttagctcttccagctttttccc 657
 Db 7156 CAATCTCCTCACTCCAACTCCACATTTCTTAAGCTGTTCACCTCCCGCTATTATCC 7215
 QY 558 cgtctat---aatctcaagaacccctgcaatcccaatctctactatagcagcgagc 714
 Db 7216 TGTCGATTCGAGCATCAGATCGGAAAGAGAAATAATCTCTCCGATATCTACTCGAGC 7275
 QY 715 ttctactctaaatatacaaaacaaatggataaaactctcaaaaacccctcaacta 774
 Db 7276 TTTTCTTAAAGAGAGAAAGAACCGCAAGATGGTTAAAGCTAGAGAGAGCGCTTACTA 7335
 QY 775 ctataatcaagtcaggtggaaactaaaacgattacgattcaactcattcccgatgcaaa 834
 Db 7336 CTATAATAAGACAGGAGTAGCTGTACAGGAGATCTGTATACACATCAITCTCTGATCAACA 7395
 QY 835 cacagcagcaaaactatttaacagggaaactcaattggcgaagcactctctggggaga 894
 Db 7396 AACTGCTCTCTGCTTTATTCACCAAGGAAGCTAGATTTGGCAAGTCTCCCTTGGGGACA 7455
 QY 895 aqcatctctcaagaacccctcaacttaacgtactgaagggcactcaactcttttga 954
 Db 7456 TAGTATTCGCAAGAACTTTACCCACAAACAAACAGACGAGCTCCCGGATCATTTGA 7515
 QY 955 tgcgcaggaacctctgctcacttcaatcaatcaataaattccctcaacaatatga 1014
 Db 7516 TATATCTGCTACTCTCTGGCTTACATTTAACTGTCGCAAAAGCTTTTAGTCATTCCAA 7575
 QY 1015 gcttagaagaaccttagcatcagccttagataaaggaagctcttctcactatattctt 1074
 Db 7576 GCTTCGCCAAAGCTTTGAGTCTAGTTTAAACAAAGAAAGCTCTTGCTC-----CTT 7626
 QY 1075 aggcgcgtgcaaaaactccgcgcatctcactcactcaaatattoatagctatccgaaca 1134
 Db 7627 GCCTTTGTTAAACCTGCAAAACATCTCTCTCGCATTTGTCACACCTTACCCAGACA 7686
 QY 1135 tcaaaaaacagagatgscacacgcgaagcttacgctaaataaactcttlaaagaagcttt 1194

RESULT 7
 AF002215/c
 LOCUS Chlamydomydia pneumoniae AR39, section 43 of 94 of the complete genome. BCT 30-MAY-2000
 DEFINITION Chlamydomydia pneumoniae AR39
 ACCESSION AF002215 AF002161
 VERSION AF002215.2 GI:8163458
 KEYWORDS Chlamydomydia pneumoniae AR39
 SOURCE Chlamydomydia pneumoniae AR39
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.
 REFERENCE 1 (bases 1 to 12980)
 AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
 TITLE Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39
 JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
 MEDLINE 20150255
 PUBMED 10684935
 REFERENCE 2 (bases 1 to 12980)
 AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 COMMENT On Jun 1, 2000 this sequence version replaced gi:7189472.
 FEATURES Location/Qualifiers
 1. .12980


```

/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(157. .1609)
/gene="CP0560"
complement(157. .1609)
/gene="CP0560"
/note="sodium:sulfate symporter family protein; identified by match to PFAM protein family HMW PF00939"
/codon_start=1
/transl_table=11
/product="sodium:sulfate symporter family protein"
/protein_id="AAF73687.1"
/db_xref="GI:8163459"
/translation="MNKKRFLSLFLTAVALGTFSPHSPASINSANOLFAIETTTI
MGIIFQVPMGAIIGISTLLITLQGLSGFNPANLVELSFSIAKGLIKTG
LGERIAYFVSALGSPGLSGVLVIDDFLAPAFSAVSARGGLIYPVVLSLDSFG
SSAEKGQDLIGSLIKVAYQSVITTSAMEFLTAMGNPLVAALAGHGVSLVIMAK
AAIIPGSLFLMFIILYKLYPPKITSCEAIRSAKRLKEMGPLKKEKTILMIFEL
LVLIWTFDGLISATTAALIGLSLLILNLDQKDVIANETAWETIWTGALIMMA
SFLNQLGFIPLVGDSSAALVGLSMKIGFLLFIYFVSHYLFASNTAHIGAPYFI
AVSLGTNPFAALTAFASNLFGGLTHYGGPAPLYFGSHLVVQEWWSGFAISI
VNIVITIGISLWAKALGI"
complement(1653. .2408)
/gene="CP0561"
complement(1653. .2408)
/gene="CP0561"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38380.1"
/db_xref="GI:7189473"
/translation="MLMDISHLEDLAVDRGILPREAIAIVKOMQIPTVLLHLHD
ATQVRPEVNDSSYOGHLYAMYLAAQFESRALPLIKLFAFEDDTPHAGLVIED
LPRLASVNDSSLIKELIETPKINPVKAAASGLVTLVAGKIPDKDVIRYPAELL
NYRLEKQPSFAMNDIAGICTLYPGELFYPIKSAKGLVGLVDTSFISMEDVENIHEET
VESCILHCSSTELINDLEMEKWELEDPIEP"
complement(2487. .2684)
/gene="CP0562"
complement(2487. .2684)
/gene="CP0562"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38381.1"
/db_xref="GI:7189474"
/translation="MKKFFVCEELCILVATHRALLETPTIALSFFKELKTYVYRAK
DILQLHNYKGFILNLSPLCS"
complement(2812. .3000)
/gene="CP0563"
complement(2812. .3000)
/gene="CP0563"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38382.1"
/db_xref="GI:7189475"
/translation="MGYQIFRLKGLGSFSSYNPNPNPYFDEGCFVYVESQFKA
LDHGILQKQETETVRNT"
complement(3012. .3809)
/gene="CP0564"
complement(3012. .3809)
/gene="CP0564"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1

```

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38383.1"
/db_xref="GI:7189476"
/translation="MNTYTFSPTLQKSFSLLEKLDYSFFFGTQTQILVITPTNIR
LAARKGCKVSTIEKIIKILSFILLPLVIIAFILYFLHKKDFKQFICPKIVSNEDE
ALLGSRPOAVKAVREISPAFFSIPKRYOLIRIDTPDKDAPSLRPIGLIILKOLCI
DTLKQSNLFLAKREMDFIHPHEEKALEDSICSEKQDQWMSLESKLLITHEFLKILFVS
GIEQLNPGFNGENGFGYFSEISTAKIFHQHGRYGPINSSGPIMKEI"
complement(4314. .5102)
/gene="CP0565"
complement(4314. .5102)
/gene="CP0565"
/note="similar to SP:P42065 PID:677944 GB:AL009126;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AAF38384.1"
/db_xref="GI:7189477"
/translation="MTYLSIKDLSLTIRGKKILNHLNMLIKGSLYLTIVGSPSGSKS
SLAULTLDLKPRTGTTTFHMDPKIPRARKVQVIVQDIDSSNPMCSIKPISSEPLNI
IGTYSKAPQNKIYNVLVNLVPSVHLKPYKLSGGQKQRIAAKALSKRELLICD
EPLSLDPLNGSILDLQFTIKKEYONTLLFITHDMSAAYIADTIANWQGSUVEHA
CREKIFSTPKHTTTDLDLAIPIFSLISTEMEPSEEEYELQVASK"
complement(5099. .5953)
/gene="CP0566"
complement(5099. .5953)
/gene="CP0566"
/note="similar to GB:M57689 SP:P24136 GB:X56347 PID:143607
PID:580898; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AAF38385.1"
/db_xref="GI:7189478"
/translation="MDNYLLNKDITITNPKRTLIENSLQKERNMLAVGESSGS
GKTTTAKIGLPELNCILIKTGSILFEDIDITKLSPELHKIRGQKATIQNAMGSL
TPSRHKGQIIEITLROHKNKEEAYNAKMLLDVDCIPNPKYSQIPFLVSGMRQ
RVYTIALASQPKLILADEDSQAQVLRILKNXQOQKATILIFTHNLSLVK
ELMNDICILIKDGKILIEITGVEIFLSPKHYTKLLNNAVSKPIPKTSSPILKNKFPQ
LMSQGGGL"
complement(5946. .6800)
/gene="CP0567"
complement(5946. .6800)
/gene="CP0567"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF38386.1"
/db_xref="GI:7189479"
/translation="MENSSAPSRSIWKSILQNKMLVLGTTTLIIMLGAALLPWFYQ
DYEQTSKDLISPCSPRFETDTRGCMFARTLGRILSLIATLIDVCGVGLW
ATVAISGKKIDFLMMRTTEILSLRPIPIILLVIFHGLLPILIAMTTIGWIPIS
RIIYQFLLLKKNKPFVLSAKAMHASTPHILKHLNLTAPISLIFITFINAYTEA
FISPLGIGIQPQASLGTLVKEGINADYFWLFFPFLSILIALSISNFIIGEGAKTL
CLEEGSHG"
complement(6831. .7775)
/gene="CP0568"
complement(6831. .7775)
/gene="CP0568"
/note="similar to SP:P24138 GB:X56347 PID:580897
GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF38387.1"
/db_xref="GI:7189480"
/translation="MFSYIKNRILFNLLSLWVILTLFVLMKTIIPGDPFNDGCVNLS
EEVLOTIKRGYGLDKPYQOYTVLHSIAKLDGNSLVYKQKVTNIISTAFPSAIL
GLQSLFLSTGGGIALGTIAALKKKQRRIYLGASIIQISIPAFIATLLQYVAVKIP

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
Patent: WO 0202606-A 128 10-JAN-2002.

REFERENCE
1 (sites)
Ratti, G. and Grandi, G.
TITLE
Immunisation against *Chlamydia pneumoniae*
JOURNAL
Patent: WO 0202606-A 128 10-JAN-2002.

FEATURES	Chiron S.p.A. (IT)	Location/Qualifiers
Source	1. .1584	
	/organism="Chlamydoiphila pneumoniae"	
	/db_xref="taxon:83558"	
BASE COUNT	506 a 378 c 256 g 444 t	
ORIGIN		
Query Match	22.8%; Score 409.4; DB 6; Length 1584;	
Best Local Similarity	54.8%; Pred. No. 3.2e-89;	
Matches 834; Conservative	0; Mismatches 681; Indels 6; Gaps	
Qy 185	tctctacatctcggggaaactcgtattataataagagatgaaccccgcttcttagat 244	
Db		
64	TCCCAATCCTCTAAAGAAATTTTGTGGTAAATATGAGGAAATGCCACGCTCCTTGGAT 123	
Qy 245	ccaagacaagtgcgacctctctcgaatacagccttgcaacaatatctatagaggatta 304	
Db		
124	CTGTGGAATAACTCGTCTCATTTGCAGACAAACTTAATGCGTCATCTATATGAAGGACTC 183	
Qy 305	gttcaagaataaatcttttcaggaataatagagcctgctcttgcaagaactactcttt 364	
Db		
184	GTGGAAGAACAATGCCAAAATGGAGAGATTAAACACAGCCCTTGCAGAAAGCTACACATC 243	
Qy 365	tctctgcagcagcactcaactatacttitaactgaatacagcttttttggagtaatggcag 424	
Db		
244	TCCGNAACGCGGACTCGGTACACATTTTAAATCAAAAACATCTCTTTGGAGTAACGGAGAC 303	
Qy 425	cccttaacagctgaagactttatagaactcttggacaagaagtagctactcaagaagtcctca 484	
Db		
304	CCTCTGACAGCTCAAGACTTTGTCTCTCTTGGAGGAAATCTCTAAAGAGATGCGGTCC 363	
Qy 485	ggaatctatgcttttgcctctgaatccaattaaaaalgtacgaagaatccaagagggacac 544	
Db		
364	TCCGTATATCTGTATGCGTTTTTACCTATCAAAAATGCTGGCAATCTTTGATGATACT 423	
Qy 545	ctctccatagacattttggagtgcaactctcctaataatgaattacactctgttgttaacctg 604	
Db		
424	GAGTCTCCAGAAATCTAGGAGTCCGAGCTTTAGNATAGCGTCATCTCGAAATTCAGTTA 483	
Qy 605	gaatccccaactcgcattcttctttaaacttttagctctcttcagcttttttccccgttcaat 664	
Db		
484	GAAACTCCCTGGCGCATTTCTCTACATTTCTTGACTCTCCTATTTTTCCTCCGTTTCAT 543	
Qy 665	aaatctcaagaagacctgcaatccaactctctta-----cctatagcaagcgagcttctc 718	
Db		
544	GAAACTCTCGGAACATATAGCACTCTTTTGGAGAGATGCCCATTTACCTGGGTGGTGTTC 603	
Qy 719	tatctcaaaatatacaacaaaacaaatggataaaactctcaaaaaacctccaactactat 778	
Db		
604	CGCCCTGTCTCTAGAAAAGCCCTGAGACTCCATCTAGAGNAAAACCCCTATGTACCAT 663	
Qy 779	aatcaaatgcaggtggaaactaaaacgattacgatttcatttcccgatgcaaacaca 838	
Db		
664	AATAAAGCCGTGTGAACACTACATAAAATTTATTACAGTTTATCTCAACGCTTAACACT 723	
Qy 839	gcagcaaaactatttaactcagggaaaactcgaattgcgaagcactccttggggagaagcgc 898	
Db		
724	CGAGCCATTCTTCACACATATGAAGAAATAGATTGGCAGGAGACTCCTTGGGGAGAACT 783	
Qy 899	attcctcaagaanaacctctacaatttacagcttaaggggcaacttaoactctttttagatc 958	
Db		
784	ATCCCTCCAGAAAATCTCAGCTTCTTACATCAAGATGACCACCTCTTTTCTCTCCGGCG 843	
Qy 959	gcaggaaactcatggtccaccttcaataataaattccccctcaacaatatgaagctt 1018	
Db		
844	GCTTCGACTACATGGTTTACTCTTTTAAATACAAAAAAACCTTGGCAACATGCTAAATTA 903	
Qy 1019	agagaagccttagcatcagccttagataaagaagccttctgtctcaactatatcttaggc 1078	
Db		
904	CGCAAGGCATTGACGCTTGCATATAGACAAAGATATGTTAACCAGTAAGTGTATACCAAGGT 963	

QY	1079	cgtcgaaataacgctgacatcctctccctacacaaattattcagctatcccgcgaactcaa	1138
Db	964	CTTGCGAACCTACAGATCATCTACATCCCAAGACITTTATCCAGGACCTATCCCGAA	1023
QY	1139	aaacagagatggcacaacgcgaagcttacgcgtataaaactctttaagaagcttttaga	1198
Db	1024	CGAAAAAGACAAACGAAGAATTCITTTGAGGCTCAACAACCTCTTTGAAGAAGCTCTAGAC	1083
QY	1199	gaactccaaactcgtctaaagatctcgaacatcttaattctatctcttcocggttctctcg	1258
Db	1084	GAATCTTAATGACACACGCGGAAGATCTAGAAGAGAAACTTTGACTTTCTCAACCTTTCT	1143
QY	1259	tcagcgaagtcttttactagtcocaaacttatcagacacagtcagtggaaagaagcttttagggttc	1318
Db	1144	TTTTTCTTACGGAAGGATTTGCCAAATGCTTAAGAGACAATCGAAGAAGCTTTAAAAATTT	1203
QY	1319	gcataccctattgtcgaagaagtaattgcttcttccaaagcagacactctcttcagggaac	1378
Db	1204	ACTATCCCTATAGTAGGCCAAGAGATTTTTCACAATACAAAAAACTTCTCTAGAGGGGAAC	1263
QY	1379	tctctttagctacagggagatggttcgcagacttctgcatctatgcattcgcattctctaag	1438
Db	1264	TATTCCTTAACCGTAGACCAATGGACCGCAGCATTTATGATCCGATGCTTTATCTCATG	1323
QY	1439	atcttggcttatccatcaggagttctctcttatgcaatcaaccataagcacttctctgaa	1498
Db	1324	ATCTTTGGCAATCCTGGAGGAATTTCCCCCTATACCTCCACAGATTCACACTTCCAACT	1383
QY	1499	attctacaataacatagaacgaagaagatacccaaaaacgcctcggaattagtcgcaa	1558
Db	1384	CTTCTCATTAAGATGACTCAAGAACAATAAAACACCTACGGAATCAGCTTATTATTGAA	1443
QY	1559	gcttctcttacctagagaccttctaatattatcagcgcattctaccacgaagcatttcaa	1618
Db	1444	GCCCTGACTATTTAGACACTCTCAATCTCGAACCACTATGTCTATCCAAATCTCGA	1503
QY	1619	tctgctatgaataaaaaacttctaatctagagctctcaccacagaggttgtagcttc	1678
Db	1504	ATTGCTTTTGACAAAAACATTAATAACTTTAATCTTTTCTCGACGACTTCGACTTT	1563
QY	1679	cgttatgctaaggaaataag	1699
Db	1564	CGTTTATAGAAAAACATATAG	1584
RESULT	9		
LOCUS	AE002315	12173 bp	DNA linear BCT 26-MAY-2000
DEFINITION	Chlamydia muridarum, section 46 of 85 of the complete genome.		
ACCESSION	AE002315		
VERSION	AE002315.2		
KEYWORDS			
SOURCE	Chlamydia muridarum.		
ORGANISM	Chlamydia muridarum		
REFERENCE	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.		
AUTHORS	1 (bases 1 to 12173) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		
TITLE	Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39		
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)		
MEDLINE	20150255		
PUBMED	10684935		
REFERENCE	2 (bases 1 to 12173) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		
AUTHORS	Direct Submission		
TITLE			

JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9710
 COMMENT Medical Center Dr. Rockville, MD 20850, USA
 FEATURES On Jun 1, 2000 this sequence version replaced gi:7190506.
 source Location/Qualifiers
 1. .12173
 /organism="Chlamydia muridarum"
 /db_xref="taxon:83560"
 /note="synonym: Chlamydia trachomatis MoPn"
 complement(97. .792)
 /gene="TC0464"
 complement(97. .792)
 /gene="TC0464"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF39314.1"
 /db_xref="GI:7190507"
 /translation="WDSCALKNLSCEVSYRIPVIATRKREVALVILVIGTVLAAIG
 ACVAIVISGVMSIGLGLVYGVLLATGFLFALMEICRAQYGRQQRIOIQOEVLESS
 PIREATLPFIEQAVYSHEPELPLSSQEEVRCRPIILQRGSKPAETPKFIAVGSVV
 ELVKVMIGPMGRGNVNPQOTLRLWDELALFALRMGELVRLDGFCECKVLPATLGDTS
 VSKTSADRASDED"
 complement(1227. .2345)
 /gene="TC0465"
 complement(1227. .2345)
 /gene="TC0465"
 /note="similar to GB:AL009126; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="guanine tRNA-ribosyltransferase"
 /protein_id="AAF39315.1"
 /db_xref="GI:7190508"
 /translation="MALRFEILHQKSKRVRGRIETEHGICIDTPAFVPAVNGALKG
 VLDNSNPLMFCNTYHLIVHPGAEVAAMGGLHOFICRNPAILTDSGGQIPFSLAIGS
 VAEIKSGKKKGNSIIKINDEGVFKSYDGRKRLSPSEVQAKHLGADIILPL
 DELLPFHPDPAVYQSSORTYANERKSLDYHLANPGKSGSMYVGHGGTFPDORLGGCQ
 FVEDLPFGSAIGSLGKNLROIVGVDTTANLSIERPHLLIGDLPLSIWATVWFG
 IDSPDSSYPYKARHGMLTSGSLKINNOYASDLNPIEPGCCACPACSQGITSRAYLR
 HLPKHPNPAGIWNASIHNMHYMQKIMSKIREKILNDQL"
 complement(2467. .3939)
 /gene="TC0466"
 complement(2467. .3939)
 /gene="TC0466"
 /note="magnesium transporter; identified by match to TIGR
 protein family HMM TIGR00400"
 /codon_start=1
 /transl_table=11
 /product="magnesium transporter"
 /protein_id="AAF73559.1"
 /db_xref="GI:8163227"
 /translation="MGEONRNEEKLVALSGNLSKTSKSHLDLSFKLEAFDTLT
 AGHSODLTYSVAYNPIDLAYVSCPLSDRSILYKSLDSIASKIAPINTDSASRW
 ALPNSLDEICALLQMPPEIWLVDIPDRRYRLDLDVKKALKRLDQKGR
 NTAGRMTNEEFAFLMPTTVEKATCIRNPNNGVDLTRLVLDKFKLQEGVFDRLSL
 IASPENLQIMPVPEHKVLAIDTTREEVVDLVERKVAIVPVEENFLGATYEDSLV
 VETIEDIADETARMAGTTEDGVHDCHVQVRELLRAPWLLITLCAGLVSASWAYFQ
 KIAPTLAMVIFPIFVNLGSLNMGVOCSTILVRSNATGTLSPGRRRETILKEMSGIL
 LTGVALGILGVCCACGLGLFTGGVQVGVTVSVGILGASLTATLGLVLSPPFF
 AKIGVDPALASGPVITLNDINSMVIFLLITGLNLVLPFK"
 4208. .4366
 /gene="TC0467"
 4208. .4366
 /gene="TC0467"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF39316.1"
 /db_xref="GI:7190509"
 /translation="MOKAHTVHPFQOLQLYDEDFTVLQAKSLGLLKTVAHYAKLAAS
 QPSQSGPI"
 complement(4328. .5410)
 /gene="TC0468"
 complement(4328. .5410)
 /gene="TC0468"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF39317.1"
 /db_xref="GI:7190510"
 /translation="MFLNLPSEVRLRPVTTESFPRASSWKATQKRITSCLTLPFPYK
 TSLTRIVCFVYGILIVLGAICRCIAYLANAGLLFSPKILAIALLSSVLLGGGLVCK
 IARKIDVIGEKIQPFAIRKWEQVITLCEKRGQTIPIODPDMYDVSLLKQSGSIAP
 VYTPPAKARTAICTACILLPITVTVLYNAFRELLIPFIVFOMIRQLYQEDLPF
 EEOPIGSDIFRMSRSFVQAVKAPFYGVACYLASLGLLNLPLUGSRVIMASVERDND
 VLRSGIWIIFCEKNCLEGGTSGGLGQHWAYLLGCFQPFRLFLKDKGEIVSGRPS
 IQAPPENKELISYLYGAALGRUTSR"
 5581. .5946
 /gene="TC0469"
 5581. .5946
 /gene="TC0469"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF39318.1"
 /db_xref="GI:7190511"
 /translation="MNYLFSLNQKLKARMLTIVTNSRIFFPFSNERILPGSTLSTD
 STFSYKILIPILLAGSICLLGALLAGTITTPAIAASYELSLGLVTVVAGIGLCSV
 FRPHLFSINQSDLIHYY"
 6012. .7034
 /gene="TC0470"
 6012. .7034
 /gene="TC0470"
 /note="O-sialoglycoprotein endopeptidase; identified by
 match to PFAM protein family HMM PF00814"
 /codon_start=1
 /transl_table=11
 /product="O-sialoglycoprotein endopeptidase"
 /protein_id="AAF73560.1"
 /db_xref="GI:8163228"
 /translation="MLTFLSSCDETSCALVENGKILANRIASQDIHAAYGVIPPEL
 ASRAHLQFLPKLAADAEVSLSDVELISVANTPGILGALSVGVNFAKGLASGLKK
 TLIGVNHVYEAHLAAACLEPSPIRFPALGLAISGHTSLIFLMPNATFLIKTRDDAI
 GETFDKARFLGLPYPGOKLEELAQDGDDEAYPSRAKYSGNDTSFSLGKTAVLIYAL
 KGNSSAKAPPEVSETQKNTIAASFQAAFWTIAQLPDIVKAFSCSELIVGGGVAN
 NRYPRLLNQCSLPTFFYSSQLCSDNAMIAGLGERLFCNTYVSKEVIPCARYQWE
 SACLSH"
 6992. .8554
 /gene="TC0471"
 6992. .8554
 /gene="TC0471"
 /note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
 PID:40005; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="peptide ABC transporter, periplasmic
 peptide-binding protein, putative"
 /protein_id="AAF39319.1"
 /db_xref="GI:7190512"
 /translation="MRKVSVGIOLLIATAITCCKSSKSNKSNHSSNOSVSVSKMD
 DPRTFPRVRLSDINLIHHLYEGLVQETPPSGEVPFPALESFELSDKTYTFHLKK
 ALMSNGDLITAHDFVSRWVNDVONRIASIVSFAFLPIDLSDKCGFFAKDNHTLVINH
 TPTPFLKLLTLTPFYPVPHBQIRNAKALPISGACFLCKKDRWLKLNKPYYY
 NKEQVALQRIHIVIPDQQTASALFKQGLDWOGUPWGHSPQEPALATANKRTPQSF
 DISGTQWLFNTSKPFFSPHKLQALSLNKEALASPTFYKPAKHLPTHLTYPEQ
 PTYKQAEVILLAKTLQALDNTMTIKOLEKPLIFSATSVSNVSNMAOMRDOWERI
 LGTIFPICGKEYALLNDLNTTFMFSINGWADFSDPLAFLSVSSKGLKPYALQDP
 LFTQLILSTETEKDPRKRI SLISBASLYLEKQNIIEPLYHDVYFAANKLSFVRLHP

gene
SCLVDMRYAKNS"
8668...9612
/gene="TC0472"
CDS
8668...9612
/gene="TC0472"
/note="similar to SP:P24138 GB:X56347 PID:580897
GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF39320.1"
/db_xref="GI:7190513"
/translation="MLGYIKKRLNLLSVVVLTFFIIRIPDPDPDENGVL
BELIALINKRGIDKDFQYLIYLKLLTLDGESLIYKDRIVIGIITLPSAIL
GLESICLALFGGITGLIAAFYKRCGRITIFFSIIQISVPFVIGAFLOYIFAIKYS
LLPTKWDGDSHLLPESIALAPMAFITOLTVASVSASLKDYLLAVANGLSPLK
LIRKVLDFPVSISYAPLITLMTGTFSIENLCIPGLGKWFICISKQRDYPMTLG
LSVFGAFPMILISLLCDLLQWIDPQIRYSYGRHSKQ"
9618...10481
/gene="TC0473"
9618...10481
/gene="TC0473"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF39321.1"
Query Match 21.9%; Score 393.8; DB 1; Length 12173;
Best Local Similarity 55.4%; Pred. No. 1.7e-85;
Matches 942; Conservative 0; Mismatches 702; Indels 57; Gaps 7;
QY 1 actccccctgctaactatgtctagataatgtctgtattgattgcaggcttagggggag 60
Db 6889 ATTTCCTTCTCCGAGTTGCTGCTGATATGCTGCAATGATAGCTGGATTAGGAGAGA 6948
QY 61 aaaaatttcaaaaaactctagattcc---ggaatttcgtatgtatgcgaagatcag 117
Db 6949 GACTGTTTGCAGTACACATATGTTCTTAAGAGAGGTTATTCCATGCGCAAGGTATCACT 7008
QY 118 gggaaatctg-----tatcaccattctcttagcctctccgttagctctcgaaggcga 171
Db 7009 GGGAAATCGCTTGCTTATCGATTAGCAACAGCAATTAACCGGCTGCTCCAAATCTCCTC 7068
QY 172 ggaagtcagtcactctctacatctcggggagaactcgtattatataagaagatgaacc 231
Db 7069 AAACAATCAAAATCACTCTCTCTTAATCAAAAGCGTATCTGTGAGCATGAAGGACGATCC 7128
QY 232 ccgtcttttagatccaagacaagtcgactctcttcagaatacagccttgctcaaacatat 291
Db 7129 TCGCAGGTTTGACCCCTCGAGAGTTAGGCTTCTTTCGGATATTAATTTGATTACCATCT 7188
QY 292 ctatgaggattagtttcaagaaaaataattcttcaggaaatatagaccctgctcttgaga 351
Db 7189 TTATGAGGAGACTGGTTCAGG---AGACTCTTCCGGAGAGGTTTCCCTGCGATTAGCAGA 7245
QY 352 agactactctcttcctcgagcagactcaattatactttttaaactgaatcagctttttg 411
Db 7246 AAGTTTCTCTATCTGAAGATTAAGAAACCTACACTTTCACCTAAAAAAGACACTCG 7305
QY 412 gagtaatggcagcccttaacagctgaagactttatagaattcttggaaacaagtagctac 471
Db 7306 GAGTAATGGAGACTGATTACTTGCTCATGATTTTGTACGTTCTTGGAGCATGTTTTCGA 7365
QY 472 tcaagagctcaggaatctatgttttgccttgaatccaattaaataaaagtacgaagat 531
Db 7366 AAATCGCATTCGAAGTATCTACTCTTTCGGCTTTTACCATTGAT----- 7411
QY 532 ccaagagggacacctctccatagaccatttttgagtgacactctctaatgaatcacact 591
Db 7412 -----TTAAGTAAAGATTGTGGATTCTTCGCTAAGGACATATACACT 7455

QY 592 tgtgttaccctggaatcccaacctcgcatttctttaaacttttagctcttccagttctt 651
Db 7456 TGTATCAATTTGCATACCCCTACTCCACACTTTCTGAAATTTACTTACCCTTCCAGTATT 7515
QY 652 ttcccccgttccata---aatctcaagaacccctgcaatcccaaatctctctacctatagaag 708
Db 7516 TTATCTGTGTCACCCGACACATCAATTCGGAATGAAGCAAAAGCTCTTCCAATATCTAC 7575
QY 709 cggagctttctacttaataataataataataataataataataataataataataata 768
Db 7576 AGGGGCTTTTGTGTTTAAAGAGAAAAAGATCGCGCTTAAATTTAGAAAAAATCC 7635
QY 769 tcaactataatacaagtcagtggaacttaaacgattacagattcaactcattccoga 828
Db 7636 CTATTTATTAACAAGAGCAAGTTTCGATCCAGAAATCCACATACATGTTGTTCTGA 7695
QY 829 tgcacaacagcagcgaacaaacttttaacagggaacaaacttaattggcgaaggaacctctg 888
Db 7696 TCAGCAGACCGCTTCTGCTTTGTTCAACAAGGAAATTTGGATTGGCAAGGACTTCCTTG 7755
QY 889 gggagaacgattctccaagaacccctatccaatttacagttctaaggggcacttacactc 948
Db 7756 GGGACACAGCATTTCCCAAGAAAGCTTTAGCAACAGCAAAATAAACGTCGTACTCTCAATC 7815
QY 949 ttttgatgtcgcaggaacccctcatggtcaccctcaatataataataataataataata 1008
Db 7816 ATTTGATATATCTGGCACCCTCGTGGCTCAGTTCAACACATCAAAAGTACCTTCAGCCA 7875
QY 1009 tatgaagcttagagaagccttagcatcagccttagtaagaagccttctgtctcaactat 1068
Db 7876 TCCCAAGTTTCGACAAAGCTTTAAAGCTTTTCTTAATAAGGAAGCTTTAGCCTCTCAAC 7935
QY 1069 attcttagcctgcgaacaaactcgcgtacatctctctacataataataataataataata 1128
Db 7936 CTT-----TGTTAGCCGCGTAAACACCTTACTTCTACACACCTGACACTTATCC 7986
QY 1129 cgacatcaaaaacagagatggcacacagccagcttaagcgaagccttaagaa 1188
Db 7987 AGAGCAACCTACTTACAACAAACAGAGCGCTGATTTTAGCCAAAATCTGTGTACAAGA 8046
QY 1189 agotttagaagaactccaaatcactgctaaagatctcgaaactcttaactctctcttcc 1248
Db 8047 AGCATTTAACAGATCTTAATATGACTATAAAGATCTTGAAAAATGCCCTTTGATTTTTC 8106
QY 1249 cgttctcctgcgcagcaagcttcttactagtcacacttaacagaaagtggaagaag 1308
Db 8107 AGGCACATCTACGTCACACTCACAATATGCCCCAAATGTATGCGCATCAATGGCGAAGAT 8166
QY 1309 tttagggttcgctatccctattgctcgaaaggaatttgccttcttccaagcagacactatc 1368
Db 8167 TTTAGGAATAACATTCCTTATTTTCGGGAAAGAGTATGCTTTTATACAAATGATTTGAC 8226
QY 1369 ttaagggaactctcttagctacagaggatgggttcgcagacttttgcctatgctatgac 1428
Db 8227 CACCAACACTTTTTTATGTCCATAAACGCTTGGTTTCAGACTTTTCCGATCCTTTAGC 8286
QY 1429 attcttaacgactctcttcttaccatcagaggattcctccttatgcaatcaacoataagga 1488
Db 8287 CTCTCTGCTGTTTTTCTT---CTAAGGGAATCAAACTTATGCTTTACAAGACCTCT 8343
QY 1489 ctcttagaattctcaaaaacatagaacaagacagacacacacacacacacacacacac 1548
Db 8344 GTTTGACCACTTATCTCTATAGAAAGGAAAGATCTCTCGAAAGCGTATTTCTCT 8403
QY 1549 agtgcgaagctctcttcttaactagagacctttctattattgagccgatctacacga 1608
Db 8404 GATTTCGCAAGCTTCTATATACCTAGAAAAACAAAATATCATAGAACCTCTATACAGA 8463
QY 1609 cgcatttcaatttgctatgaataaaaaactttcttaactctagggagctcaccacagaggt 1668
Db 8464 TGTTTTTCACTATCGCGCAAAATAAATAAACTTTCTTTTGTAGATTACATCCCTCAGGCT 8523
QY 1669 tgtggacttcgcttatgctaa 1689

Db 882 CCTTCTTATATACCTGCAAAACCTCTAATACAAATTAAGCCCTCAGGAAGCCAT 941
 QY 1030 agcatcagccttagatagaagaagctcttgcttgcacactatattcttaggcgctgcacaaac 1089
 Db 942 TGCTCATGCTATTGATGAGAAATCTATCTTAAGACTCGTGCCTTCAGGACAGAA----- 996
 QY 1090 tgcgcatcctctcactacacataattcatagctatcccgaaacatcaaaacaaagagat 1149
 Db 997 -GCTGTAACTCTAGTTCCTCCCAATCTTTCACAACTCAATCTTCAAAAGAGATCTCAAC 1055
 QY 1150 ggcacacgcagccttagctgtaaaacactctttaagaagcttttagaagaactcccaaat 1209
 Db 1056 AGAAGAAGCAGCAACAAAGCCAGAGCATATTTTCAAGAAGCTAAGAAACAC-----T 1109
 QY 1210 cactgctaaagctcgaacacttaattattcttctccgcttccctcgcgcagaagctc 1269
 Db 1110 TTCTGAAGAAGAACTCGCAGAACTCAGACCTCTCTATCTATAGATTCTCTGAATTCCTC 1169
 QY 1270 ttactagtcacacttaccagacagctggaagaagcttttaggcttgcgctacccctat 1329
 Db 1170 CATCATAGCTCAAGAAATCCAAAGCAACTTAAGATACCTTAGGATGAAATCAAAAT 1229
 QY 1330 tgcggaaggaattgtctctccagacagacctattcttcaggaaactctcttttgc 1389
 Db 1230 CCAAGCAGTGGAGTACCACCTGCTTTTAAAGAAACGTCGCAAGGAGATTCTTCATAGC 1289
 QY 1390 tacagagagatggttcgcagactgtgctgactgactggaacttcttaacagatctttgttga 1449
 Db 1290 GACAGAGGATGATGCGGAATACGTAAGCCCGTAGCTTCTCTATCTATCTTAGGCAA 1349
 QY 1450 tccatcagaggtctctcttcttgatgaatcaaccataagga 1488
 Db 1350 CCCCAGAGACCTCACACAATGGAGAACAGTGTATACGA 1388

RESULT 12
 AX300389
 LOCUS
 DEFINITION Sequence 1 from Patent WO0185972.
 ACCESSION AX300389
 VERSION AX300389.1 GI:17381780
 KEYWORDS
 SOURCE Chlamydomophila pneumoniae.
 ORGANISM Chlamydomophila pneumoniae
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE
 1 (sites)
 Mardin, A.D., Oomen, R.P., Wang, J. and Dunn, P.
 Chlamydia antigens and corresponding dna fragments and uses thereof
 Patent: WO 0185972-A 1 15-NOV-2001;
 Aventis Pasteur Limited (CA)
 FEATURES
 Location/Qualifiers
 1..1787
 /organism="Chlamydomophila pneumoniae"
 /db_xref="taxon:83558"
 101..1687
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD13017.1"
 /db_xref="GI:17381781"
 /translation="MKMRLKPTLKSILPNLLFLTLSSCKQKPELQKHLVIAMS
 HDLADPNVILSRDASLAKALVEGLTRETDOGLALALAESYTLKSDHKVYIFKLKP
 SVNSDGTPLTAVDFEKSILKLYPEESPSIHLILGVINKNSAIHNAKSLFTLGLQAK
 DDLTVITLQEPYFYLTLARVFSVPHVTLRESYKKGTPPSTYSNGPFVLKKEH
 QNVLLKRNPHYDHESSKLDVTLKILIPDASTATKFKSIDWIGSPWSPAINSD
 QVLSQELIYSSSTLLIYNLQPLNKLAKLAIAHIDRKSILRLVPSGPEAV
 TLVPLNSQLMKEISTEERTOKARAYFOEAKETLSKELAEILILYIDSSNSII
 AOEIQRLKLTGLIKIKQMEVHCFLKRRGRODFIATGWTAEVSPVAFSLIGN
 PDELQWNSDYETLKLKLYPHAYKENLKRAMEIIEETPIIPLYHGKVIYAIHPKI
 QNTFSLIGHTDLKNIDLS"
 628 a 436 c 261 g 462 t

BASE COUNT
 ORIGIN

Query Match 10.1%; Score 181.4; DB 6; Length 1787;
 Best Local Similarity 49.3%; Pred. No. 1.1e-33;
 Matches 640; Conservative 0; Mismatches 626; Indels 33; Gaps 5;

QY 202 agaatcctctatttaataagaagatgaaccccgctttcttagatccaagacaagtgcgact 261
 Db 211 ACATCTCGTTATTTCGATGAGCCATGATCTCGCGACCTAGATCCTCGCAATGCCTATT 270
 QY 262 tctttcagaatacagctctgcaacatatctatgagggattagttcagaagaataatct 321
 Db 271 AAGCAGAGATCTTCCCTAGCAAAAGCCCTCTATGAAGACTGCAAGAAGAACTGATC- 329
 QY 322 ttcaggaataatagagcctgctcttgcaagaactactctcttctcgcgcagactcac 381
 Db 330 -----AAGGAATCGCACTGGCTCTTGCAAGAAGTTATACCTCTGCAAAAGATCATPAAGGT 384
 QY 382 ttatacttttaaaactgaatcagctttttggagtaattggcgaccccttaacagctgaaga 441
 Db 385 CTATACCTTTAAACTCAGACCTCTCTGTGTGGAGCGATGGCACTCCACTCAGTCTGCTATGA 444
 QY 442 ctitagaatcttggaaacaagtagctactcaagaagctctcaggaatctatgctttgc 501
 Db 445 CTTTGAATAATCTATAAAACAACTGTACTTCGAAGAAATTTTCACCTTCCATACATACTTT 504
 QY 502 ctggaatcccaattaaaaatgtacgaagatccgaagaggaacacacctctccatagaccattt 561
 Db 505 ACTCGCGTGTATTAANAATCTTGGCAATCCCAATGCTCAAAATCTCTGGAATCTCT 564
 QY 562 tggagtgacactctcctaatgaatctacactgtttgttaccctgggaatcccaacctcgca 621
 Db 565 TGGGATACAGGCAAAAGATGATCTTACTTTGGTGATTAACCTAGAGCAAACTTTCCCAT 624
 QY 622 ttcttaaaacttttagctctccagctctttttcccggttcaata-----aatc 669
 Db 625 CTTTCTCACACTTATGCTGCGCCGCTATTCTCCTCTGTTCATCACACCTTAGGGAATC 684
 QY 670 tcaagaacctctgcaatccaaatctctacctatagcaagggaggtcttctcactaataaa 729
 Db 685 CTATAGAAGAGGACACCCCACTCCCATACATCTCCAATGGGCCCTTTGTCTTAAAAA 744
 QY 730 tatcaacaaaaacatggataaaactctcaaaaaacctactactataataaagtc 789
 Db 745 ACATGAACACCAAACTACTTAATTTTAGAAAAAATCTCTACTACTATGATCATGAATC 804
 QY 790 ggtlgaactaaacagattagactcactctcccgatgcaaacacagcagcaaaact 849
 Db 805 AGTAAAGTTAGACCGAGTCACTTTAAAAAATTTATCCAGACGCTCCACAGCCAGAACT 864
 QY 850 attaatcagggaacactcaattggcaaggacacctctctgggagaagcattcctcaaga 909
 Db 865 TTTCAAAAGTAATCTATAGATTGGATTGCTCACCTTGGAGCGCTCCGATATCTAACGA 924
 QY 910 aacctatccatttaacagctaggggcacttacactcttttgatgtcgcaggaacctc 969
 Db 925 AGACCAA---AAAGTTCTCTCCCAAGAAAAAGATTCTTACCTATTCTGTTTCAGGCCAC 981
 QY 970 atggctcacttcaataataaaatcccccctcaacaataatgaagcttagagaagcctt 1029
 Db 982 CCTTCTATCTATACCTTGCAAAACCTCTAATACAAATAAAGCCCTCAGGAAGCCAT 1041
 QY 1030 agcatcagccttagatagaagaagctcttctcaactatattttaggcgcgcagcaaac 1089
 Db 1042 TGCTCATGCTTATGATGAAATCTATCTTAAGACTCGTGCCTTCAGGACAGAA----- 1096
 QY 1090 tgcgcatcctctcactacataatctcagctatcccgaaacatacaaaacaaagagat 1149
 Db 1097 -GCTGTAACCTAGTTCCTCCCAATCTTTCACAACTCAATCTTCAAAAGAGATCTCAAC 1155
 QY 1150 ggcacacgcagccttagcctaaaaaacctctttaagaagccttttagaagaactccaat 1209
 Db 1156 AGAAGACGACAAACAAAGCCAGAGCATATTTTCAAGAGCTAAAGAAACAC-----T 1209

product: hypodermal protein
protein_id="AAF39298.1"
/db_xref="GI:7190489"

product: hypodermal protein
protein_id="AAF39298.1"
/db_xref="GI:7190489"

```
/protein_id="AAF39298.1"
/db_xref="GI:7190489"
```

```
/protein_id="AAF39298.1"
/db_xref="GI:7190489"
```

product: hypodermal protein
protein_id="AAF39298.1"
/db_xref="GI:7190489"


```

/translation="MFISDKTSLSRHSTDLAEAFGNKLLYPVTOGFGTRISVKPVIV
DDMTIHGFYSKQSLCKRICCVIACILLAPVLLMTLIGVIANVCSATYKRVQDLYLA
TGLEFGKEMGPWTCAPI"
gene
complement(6896..7135)
/genes="TC0445"
CDS
complement(6896..7135)
/genes="TC0445"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39299.1"
/db_xref="GI:7190490"
/translation="MGANKIHAIQTQIRLQDSCFEKIQNSSLOPLNVIALIFELN
PNAEVTGLRSLFLPIISAASARVPLNEMIFSPKE"
gene
7303..8886
/genes="TC0446"
CDS
7303..8886
/genes="TC0446"
/notes="similar to SP:P26906 PID:48808 GB:AL009136;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein"
/protein_id="AAF39300.1"
/db_xref="GI:7190491"
/translation="MPKKFLYLIPFLSTFGLTSHOCKEENLRILRVAICHDPMS
LDPROVFLKDVCIKALYEGIVRENDGSLHLALARYLSLSDRCVYFFLKKTFWHN
GLDVTAYDFEESIKOYFLEVDNVALRLILKNSHAVLKGDIPVENLGVRAIDHETL
EITLESFSSFLTLTHFVFPVHASLREYIRNRSKSLPIISNGPFIIRCYEPQNL
LQKPLNPDHQNVDLAVRQIYVDIHTAVQLFQKYYDVLGLPLWSPFPLEQKVL
SDPLNDYVLANCTVFCNVNHPKLDNPSLRALSIAIDRETLKLACKGSIATSPVH
PSLSKPLDVLQSKRLEISLAKNYLAELKTVPOBELKKTILYPIESVTLVRAVOEIR
QQLFDVLGPKISTGLEVHSLFKRSKEPSLSTGNWADYQQAFAFSLGNGTKYK
DYOVDMQOEYTDIVSRLLVEDSDTDLQILAEOLLESFLPIPLYHLDYATAKHPKVS
NLQTSLSGEIDLKRVSLVE"
gene
9068..10144
/genes="TC0447"
CDS
9068..10144
/genes="TC0447"
/notes="similar to GP:2444077; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phospholipase D family protein"
/protein_id="AAF39301.1"
/db_xref="GI:7190492"
/translation="WTSPLSTAASSYLRTLQRAFPLGGGGYTPNPSAQTALRVQTA
TSLAIVPVPPLSSPVSEFSKCKNDQAQIENAILSSASSVFLKIFSLSSPIVD
LICKSESIPVTHIQHPEPIYVKNLEAGAEILRCHNRKSLRLKRTMILDEKQVIT
GSANFRNSINDVLLARVNVHIANLQKQKQKGAIIENSTKGTVCYLPIINHKK
CGNEROIVKAIINNTSSQIGMCIITHRGILQALNEATQSVLVTIIDSLSQOIT
DIILKALSKLRVVGTDRIHCKACILDHNTAIGSANWAASGLKANKEDIIVNPLT
ERODELSIWRYLCDNSAILED"
gene
complement(10175..10582)
/genes="TC0448"
CDS
complement(10175..10582)
/genes="TC0448"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39302.1"
/db_xref="GI:7190493"
/translation="MKILRSYCYLFAWLVSIGITLMSVYVYLLNVEPCVLYYORI
CLPPLVVIIGISAYLDDLSVKIYALPLALGICIAIYQVQLQEIPLGMITDLCGVSCS
TKLFLLGFTTPMASALFAFNAIFATKSE"
gene
complement(10382..11295)
/genes="TC0449"
CDS
complement(10382..11295)

```

```

/genes="TC0449"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
Query Match 9.0%; Score 161.6; DB 1; Length 11402;
Best Local Similarity 48.0%; Pred. No. 6.3e-29;
Matches 711; Conservative 0; Mismatches 729; Indels 42; Gaps 7;

QY 211 tattaataataagagatgaaccccgcttcttttagatccaagaagaagtcgactctttcaga 270
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7407 TGTTCGTAATTGGCAGCATCCCAATGCTTTAGATCCTCCCAAGTTTTTTTGTGATTAAAGA 7466
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 271 aatcagcccttgtcaaacatatctatgaagagatagttcaagaagaaataatctttcagaaa 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7467 TGTTCGTAATGACAAAACACACTTTATGAAGGATGGTTTCGAGAAAAATGAT-----GSCAG 7520
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 331 tatagagccctgtcttcgagaagactactctcttctcttcgagcgactcaactatacttt 390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7521 TCTTCATTTGGCTCTAGCAGAAAGATACAGCTTGTCCAGGATCGGTGTGTCTACACTTT 7580
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 391 taaactgaatacagcttttttgagtaatgagcgaccccttaacagctgaagactttataga 450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7581 TTTTCTAAAAAAAACCTTTTGGCATACGGAGATCTTGTAAACAGCATATGATTTTGAAGA 7640
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 451 atcttggaacaaagtagctactcaagaagctcaggaatctatgcttttgcttgaatcc 510
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7641 GTCTATTAAGCAATTTTATCTCAATGAGGTGATAATGTGTGGCTTACGATTCGTGGGTT 7700
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 511 aattaaaaattctacgaagatcccaagaggacacctctccatagaccattttgagtgca 570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7701 AATTAAAAATCTCATGCTGCTTCTTAAAGGAGACATCTCTGTTGAAATTTAGTGTGTG 7760
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 571 ctctcctaataatcactacattgttgaatccctgggaatcccaacccctgcattcttctaa 630
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7761 CGCTTTAGATGAACATACTCTTGAATCACTCTAGAACACCCCTCTCTCTCATTTTTTTAGA 7820
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 631 acttttagctcttcagctcttttcccgcttcaataatctcaagaacccctgcaatcaa 690
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7821 AACTCTCACACATCCGGTTTTTATCCAGTGCATGCTTCTTTGAGAGAAATATTATAGAAA 7880
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 691 atc-----tctacctatagcaagcgagcttctctatccctaaaaataca 735
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7881 TCGAAGTAAACGCTCTCTCCCAATAATTTCTAACGGCCCTTTTATCATTCGGTGTATGA 7940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 736 acaaaacaaatgataaaactctcaaaaacccctcactactataatcaaaacaggtgga 795
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7941 GCCACAAAATTTTTTGTGCTGTGATAAAAATCCTTTTATCATGACCAAAAAATGTTTC 8000
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 796 aactaaacgattacgattcattcccgatgcaaacacacagcagcaaacattattaa 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8001 TTTAGATGCTGTACGCTTACAAATAGTTCGCGATATTCATACAGCGGTGCAATTTGTTCA 8060
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 856 tcagggaacaaactcaattggcagacccctctgggagaaagcattctcctaagaacacct 915
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8061 AAAAAAATATGTTGATTTAGTGGGGTTGCCCTTGGAGCTCCTCTCCCTCTAGAAAGAAC 8120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 916 atccaatctacagtcctaaagggcacttctcttcttgatgtcgcaggaaacctaatgct 975
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8121 AAAAAATCTT---TCTCAGATTTTTTTGATGATTATCCGGTGTAAATTCGACAGTTCT 8177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 976 caacttcaataataaattcccccctcaacaatatgaagcttagagaagccttagcatc 1035
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8178 ATTCTGCAATGTTACCCACAAACCAATAGATATCTCTTCGCTAAGACGGCGCATCTCTTT 8237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1036 agccttagtaaggagcctctgtctcaactatattcttagccgtgcacaaactccga 1095
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8238 AGCAATTGATCGAGAAAC-----ATTATTAATAATAGCTGCAAGAAAGCATTCCTAC 8291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1096 tcatctctcctacataaattatcagctatcccgaacatcaaaaaaagaagatgacaa 1155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8292 CAGTTTGTACATCTAGTCTTCTAAGATGCTTTTAGATGTTCTTCGAAAAAAGAGAG 8351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

/strain="D/UW-3/CX"
 /db_xref="taxon:813"
 206..1384
 /gene="trpB"
 /note="CT170"
 206..1384
 /gene="trpB"
 /codon_start=1
 /transl_table=11
 /product="Tryptophan Synthase (Beta Chain)"
 /protein_id="AAC67761.1"
 /db_xref="GI:3328574"
 /translation="MKKHPGGGAFPEELLAPIONKAEWEILKTOQSFSLSELD
 LKNVAGROTLETVKNFARADIPRVFLKREDLLHTGAHLNNALGOCILLAKYLGKT
 VVATGAGGVATACAYALGLDCCVVMGAKDVERQKNPKVRFGLGAEVYSYVK
 CGLADAVNOALOWATHTSFHYICLGSALGPLPDPIVRFQCSVISAEVKEQIHAVA
 RDPIILACIIGGSNAIGFFPHFIPNPKVILIGVEGGGLISGSKHGAARFATGRGV
 HGFYSYLLQDDQCVLQHTSISAGLDYPSVDPDHMHESGRAFYTLATDEALRAFI
 LITNNEGIPALLESSHALAHLVSIAPSLPKEQIVIVNLSGRGDKDLQIIRNRGVIY
 1377..2138
 /gene="trpA"
 /note="CT171"
 1377..2138
 /gene="trpA"
 /codon_start=1
 /transl_table=11
 /product="Tryptophan Synthase (Alpha Chain)"
 /protein_id="AAC67762.1"
 /db_xref="GI:3328575"
 /translation="MSKLTQVFKTKLCIGYLTAGDGGTSYTEAAKALIQGGVDIL
 LGPFSDPADNPEIQVSHORALAEMLSETLLEIGIRAFNQEVPILISYVNPIL
 QRDLYLRLKDKAGIVGVVIDLPAPLSHGKSPFFEDLLAVGLDPIILISAGTPPE
 MSLIQYARGFLYIPQCATRDSEVGIKEEFKYREHFDPIDVDRDIDCKKAAHVI
 NYSDGPIKATPVHCTTMDSSVETLIALAQTVIPG"
 complement(2385..2876)
 /gene="CT172"
 complement(2385..2876)
 /gene="CT172"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAC67763.1"
 /db_xref="GI:3328576"
 /translation="MNYHTFVKTSMFFLAKLVQKNKPNFLIKKESETVLFIFERQ
 LKMWGYSIDNNYISDYNMEFGRPLLOKLANPVCKLOLEQARTLSNVTYVG
 DIVLMRSPFEKSVLLTLINEILIQESLFLFKAPENVQCPKMSFEHGAHILLKILF
 TVS"
 complement(2904..3074)
 /gene="CT172.1"
 complement(2904..3074)
 /gene="CT172.1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAC67773.1"
 /db_xref="GI:3522888"
 /translation="MLRLFCIGYFPQLEFKKNLLEDLYNKLFEKIVVMGSGACV
 NLAIGGGTILL"
 complement(3142..3414)
 /gene="CT173"
 complement(3142..3414)
 /gene="CT173"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAC67764.1"
 /db_xref="GI:3328577"
 /translation="MKICRETCDAILKARDFAFEIERTYKNGSVIYTSRQCPGVEIRE
 IGDQAKALLHNOEELQVAAILGHKLITPPUSIYKHSLLYKRLVI"
 complement(3501..3956)
 /gene="CT174"
 complement(3501..3956)

gene	GIAPILSLGRL"	complement(8249. .8761)	
CDS	/gene="CtI79"	/db_xref="GI:3328583"	
	/gene="CtI79"	/translation="MLAAGAFIALCIGCVLCCIDNVFLSEIFLPFILGILLSILLI ALLRLAKWEOKLAAQKQRIASCFSLACRTYKSGAPSVKGLVNPISQISELPSG FSRFRFAYLTQAKPSLLTIKRIQLTTPFDEITAFSHIREGLYSEIQRDHDKKL DNKANTSPKG"	
gene	/codon_start=1	complement(8905. .9597)	
	/transl_table=11	/gene="taub"	
CDS	/product="hypothetical protein"	/note="CtI80"	
	/protein_id="AAC67770.1"	complement(8905. .9597)	
	/db_xref="GI:3328583"	/gene="taub"	
	/codon_start=1	/codon_start=1	
Query Match	6.5%;	Score 117.6;	DB 1;
Best Local Similarity	46.2%;	Pred. No. 3.3e18;	Length 10827;
Matches 685;	Conservative 0;	Mismatches 754;	Indels 45; Gaps 7;
Qy	212	attaataataagatgaacccggttcttttagatccagaagtcgacttctttcagaa	271
Db	4256	ATTGCGATCTCATGATCCAAATGCTTTAGATCCGCGTCAGGTTTTTAAGCAAGAT	4315
Qy	272	atcagcccttgcacacatctatcaggagatagttcagaagaaataatctttcaggaaat	331
Db	4316	GTTTCTATTGTAAAGCTCTCTATGAAGGTTAGTCGGGAAAGAGCTGGT-----	4370
Qy	332	atagagcctgctttcagaagactactcttcttcogcagcactcacttatactttt	391
Db	4371	-TCCAGCTAGCTTTGCGACAAAGATATCATCAATCTGATGTTGGTTGTGTTTACTTTT	4429
Qy	392	aaactgaaatcagcttttttgagtaatggcagcccttaacagctgaagactttataga	451
Db	4430	TTTCTAAAAATACATCTCGAGCAGCGAGATGTTGTAACAGCATATGATTTTCAAGAG	4489
Qy	452	tcttggaaacagtagtactcagaagtcgaggaatcgtctgttgccttgaatcca	511
Db	4490	TCTATTAAACAAATTTATTTCGGAGAAATTGATAACCCCTCGTTACGCTCTCTGCATTA	4549
Qy	512	atataaatgtacgaagaatccaagaggagacacctctccatagacaaatttggagtcac	571
Db	4550	ATYAAAAATTCTCATGCTGTTTTAACGAGAGCTCTCCCTGTTGAAGATTTAGTGTTAGA	4609
Qy	572	tctctaatgaattctacactgtttgttacccttggaaatcccaacctgcatttctataaa	631
Db	4610	GCTTTGAAATCGGAAAACCTAGAAATTGTTTTAGAAAACCCGGTTTCCTTATTTTCTAGAG	4669
Qy	632	cttttagctttccagtttcttccogttcataatctcaaga-----	676
Db	4670	ATATTGGGCACCCGGTTTTTATCCGGTGCACACTCTTTACGAGAATATTACAAGAT	4729
Qy	677	---accctgcaatccaaatctctaacctatagcaagcggagcgttctctataaatc	733
Db	4730	AACCGTAACAACGGGTTTCCCGATAATTTCTAATGGTCCCTTTTCGGATTCATGTTAT	4789
Qy	734	aaacaaacaaatggataaaactctcaaaaacctcactactataataatcaagtcaggtg	793
Db	4790	GAGCCGCAAGATATTTACTTAATCAACAANAACCTCTGTATCATGCCAACGACGATGT	4849
Qy	794	gaaactaaacgattacgattcacttctcccgatgcaaacacacagcagcaaaactattt	853
Db	4850	CTGTTAAATTCGGTATGTTTGCAGATAGTTCCTCATATCCATACAGCTATCCGATTATTC	4909
Qy	854	aatcagggaacaaactcaattggcaaggaacctcttggggagaacgcattctcagaacac	913
Db	4910	CAAAAATATCATCGGTTTGGTTAGTTCGGTTACCTGGAGCTCCTCTTCTTTTGAAGAA	4969

Qy 914 ctatccaatttacagtcgttaagggcacttacactcttttgatgctgcaggaacccctcag 973
 Db 4970 CAAGAAATTCCTCC---CTAGAGAAAATTAATTGATATCTCTGATGATGCTCTGTT 5026
 Qy 974 ctacactcaatatacaataattccctcacaatatgaagcttagaagcccttagca 1033
 Db 5027 TTATTCGTACATTCATCAACACCTTTTAAATAATCCCTCGTGAGAACACCCCTCT 5086
 Qy 1034 tcagccttagaagaagcctctgtctcaactatattcttaggcctgcgagcctcactgc 1093
 Db 5087 TTAGCAATCAATCGAGAACTTTATTAAACTAGCAGGTAAGGCTGTAGCCTACGAGC 5146
 Qy 1094 gatcatctctactacaataattcaatagotactcccgaaactcaaaaacagagatgca 1153
 Db 5147 TTGTGTTACCCAC-----AATTATCTCAGATACCTGCTACTTGTCTCAAGATG--- 5198
 Qy 1154 caagcgaagcttacgctcaaaaactctttaagaagccttagaagaaccccaactcact 1213
 Db 5199 -AGCGATTGCTTTAGCAAAAGGCTACITGACCGAAGCTTTAA-----GACTTTATCT 5251
 Qy 1214 gctaaagatctcgaaacttaattcttctccgcttctccgcttctccgcttctccgct 1273
 Db 5252 CAAGAAAGATTAGAAAAATTAATTAATTTATCTATAGATCTGTTGCTTACGAGCC 5311
 Qy 1274 ctagtccaactatcacgaacacagtggaagaaagtttaggttcgctatccctattgct 1333
 Db 5312 GTTGTCAAGAAATTCGCCAACAAATATTATTTAGTACTGGGATTTAAAAATTTCTACATTA 5371
 Qy 1334 ggaaggaattgctctctccagcagacactctctccgaggaactctctttagctaca 1393
 Db 5372 GGAATAGAAATATCAATGTTTTTATAGAAAAGCTTCCAGAGGAGAAATTCCTCTAGCAACT 5431
 Qy 1394 ggaaggaattgctcagcagacttgctgactcctatgcttgcgaggaactctttagctcca 1453
 Db 5432 GGAATAGAAATATCAATGTTTTTATAGAAAAGCTTCCAGAGGAGAAATTCCTCTAGCAACT 5491
 Qy 1454 tcaggagttcctccttatgcaatcaaccataaagactcctagaaattctacaaacata 1513
 Db 5492 ACAAGATATAAAGACTTTCAATGTTTAACTGCAGAACCAAAAGTACACAAATATAGTT 5551
 Qy 1514 gaacagacgaagatcaccaaaaacgcctcggaatttagtgcgagctctcttacccta 1573
 Db 5552 GCTCAACTCTGATTAGAAATCAAGACGACCTACAGCTTATGGCAGACGATGTTGCTT 5611
 Qy 1574 gagaccttcattatttagcgcgatctaccacagcagcattcattgctatgaataa 1633
 Db 5612 AAAGAAAGTCCCTTATTT---CTCTATACCCACCTCGATTTATGTTATGGAACAGCCT 5668
 Qy 1634 aaacttotaatctaggagcttcaccacaacagaggttggaact 1677
 Db 5669 CGGGTCTGATCTCCAAACCTCTCTCTCGTGAGAAATGATTT 5712

RESULT 15
 AF305387
 LOCUS
 DEFINITION Bacillus thuringiensis 9364 bp DNA linear BCT 30-AUG-2001 sequence.
 ACCESSION AF305387
 VERSION AF305387.1 GI:10717142
 KEYWORDS
 SOURCE Bacillus thuringiensis.
 ORGANISM
 Bacillus thuringiensis
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
 1 (bases 1 to 9364)
 Gominet,M., Slamti,L., Gillois,N., Rose,M. and Lereclus,D.
 Oligopeptide permease is required for expression of the Bacillus thuringiensis plcR regulon and for virulence
 Mol. Microbiol. 40 (4), 963-975 (2001)
 JOURNAL 21295338
 MEDLINE 11401703
 PUBMED
 REFERENCE 2 (bases 1 to 9364)

AUTHORS
 TITLE
 JOURNAL

FEATURES
 source

Gominet,M., Gillois,N., Rose,M. and Lereclus,D.
 Direct Submission
 Submitted (13-SEP-2000) Biotechnologies, Institut Pasteur, 25 rue
 du Docteur Roux, Paris 75015, France
 Location/Qualifiers
 1. 9364

/organism="Bacillus thuringiensis"
 /db_xref="taxon:1428"

1249. 7088

/gene="oligopeptide permease operon"

1332. 2987

/gene="oppA"

1332. 2987

/gene="oppA"

/codon_start=1

/transl_table=11

/product="OppA"

/protein_id="AAG22034.1"

/db_xref="GI:10717143"

/translation="MKKKIPILLASTLTASTLMLGACGYSQKDDAKAKGKNTSSNGK
 QVNLDELSEIPSDASLASDASSASTALNTMEGLTRIGKDKRVPVGIADVDQKDDG
 KKYIFKLKDAKWSNGBEVTAQFVYSWRAVNPDTKATYSYIMFDIKNAEKIKKEL
 PADGLVKALDDITILEVDNPVYFDLTVPYFVPLNENFVKAQDQKGLANFTL
 YNGPFVDMKHEQSFQKKNPSYWDNKTVEEINFNIVKNTDNLVETSDRA
 LNTSEFVKFROSSEFTKREAGVAVLRFNOSNYLSNKLKALSMKSDRDNIAKYI
 ALNGAIGAYFVGDFAGPNKDFRAENGKLVETNPKEAKLWETAKKELGDKTEL
 EFLPNDEDAKVGFEKGEKNLPLSIIKIQQFPAQKNLEDSQQYDIAGFWGP
 DRDPTSYLDMFVTSNGKTYGSKYDELILKARTDKDLQARNNLLEVEKMLIK
 EDVATIPFQKGSAYIVKDAVKDILIPINGGLRTYKWASEQK"

3115. 4044

/gene="oppB"

3115. 4044

/gene="oppB"

/codon_start=1

/transl_table=11

/product="OppB"

/protein_id="AAG22035.1"

/db_xref="GI:10717144"

/translation="MGRYVLRVYVYMAVLTFLITLTFILMKLLPGSLKNOEKLSPA
 QKEITLERYGLNDPVPQVYARVYGLNKLKGLDGLSVFOYDNPVDMVIDRIGFSAQLG
 QAILIGVFIGLIGVIAALRNNTWDYCATIISVIGMSVSPFVFAALQYFVYKGLW
 FPAVFWKGPFTNPTIALSMVAIATIAFRAELIEVMOADYILTAKAGISQGVII
 VKHALRNALIPVVTILGPMVAALITGTLVIEQIYAVPGLGEQFVKSITVNDYTVIMGT
 TIFSAIFILVIFIVDILYGIIDPRIRLAGGK"

4044. 5057

/gene="oppC"

4044. 5057

/gene="oppC"

/codon_start=1

/transl_table=11

/product="OppC"

/protein_id="AAG22036.1"

/db_xref="GI:10717145"

/translation="MKDYKLSPLDFQAGNQNVDNEVIAVPSLTFWQDVRRLFOHK
 GATGIVLLSILLIILAIIPGWPVSKHSEKQDLGRAMPKPIPIENHWPFGTQY
 GVQYKEDIKYEFNFGFDDLDRLWTRTWESTVSLYALAAALDVLGVAYGGS
 AFVGRVDNIMORIMEIIPYLIIVILVINGSGIWSITLAMAITIGMGSIVR
 GQILKNOEYVLSATLGATNTQIIVKHLIPNMGPPIVMTPTIPSAVGEAFSLF
 IGLGIQFPFASISLVNDGYSIQTPHMFIPAVVISILILAFNLMDGLRDLDPK
 MRK"

5076. 6119

/gene="oppD"

5076. 6119

/gene="oppD"

/codon_start=1

/transl_table=11

/product="OppD"

/protein_id="AAG22037.1"

/db_xref="GI:10717146"

/translation="MKTLLKLVKQVSDFTHAGEVQAVRGVTFDLKGETLAIVGSSG
 SGKSVTSKALMGLIPNPGRIKNGEIVFEGDLTKLEKEMQVQRKGDIAIFQDPM
 SLNPTWTIGNOIMEGLIKHOGSKADARKVALELIDLVGIPNPEARKQYHQFSGGM
 RQRVIANALACNPKLLIADEPTALDVTIQIQIILMKDIOQKTEAAIIFITHDLGV

VANADRVAVNVAKYVEIGTVDEIFYPNPKHPYTWGLIASMPSLDGSEBEELYAIPGTP
 PDLKPKGDAFAPRNQALKIDFEMDPLEKVSQDTHYAATWLLHQEAPVAPVAVE
 KRILQMKAGEQHD"

gene

6112..7047

CDS

/gene="oppf"

6112..7047

/gene="oppf"

/odon_start=1

/transl_table=11

/product="oppf"

/protein_id="AAG22038.1"

/db_xref="GI:10717147"

/translation="MTKQREKLEIVKVKVHDFVSGVYKAVNDISFDIYRGFTFLV

GESCGSRSTGRTIILRYDATAGEVLDGENVHGKRSRAELKFRNKKMOMIFQDPYAS

LNPRTVGDIIAGSDIHGLAKSKERMDRVHLLNTVGLNKEHANRPFHFSGGQKQ

RIGTARALAVEPFIADPEISALDVSIAQVWNLKLLQKKEKGLTYLFIADLSMWK

YISRIQVMTYRGVIVSELYANPVPHYTKSLLSAIPLPDPDYERNKRKRIYVDYS

QHDYGREVPTMREIRGFHVLCSAEYKKEIYQ"

BASE COUNT 3265 a 1567 c 1818 g 2714 t

ORIGIN

Query Match 4.4%; Score 79.6; DB 1; Length 9364;
 Best Local Similarity 46.9%; Pred. No. 6.1e-09;
 Matches 352; Conservative 0; Mismatches 389; Indels 9; Gaps 3;

Qy 328 aaatatagagctctcttgcagaagactactctcttctcgcagagactacttatac 387

Db 1589 AAGAGAGCTGCTGGATTGCGAAGATGTTCAAAAGGTAGATGACGGAAGAAATACAT 1648

Qy 388 tttaaaact---gaatcagctcttttggaagtaatggcgaccccttaacagctgaagactt 444

Db 1649 ATTAAATTAAGAAAGATGCAAAATGGTCGAATGGGAGCGCTGTACAGCAAAAGACTT 1708

Qy 445 tatagaactcttggaacaaagtagctactcaagaagctctcaggaatctatgtctttgcctt 504

Db 1709 TGTGTATTTCATGGAAAGAGCAGCAATAATCCAGATACAAAAGCGACGTATTGCTACATTAT 1768

Qy 505 gaatccaattaaaaattacgaagaagatccagaagggacacacctctccatagagaccatttgg 564

Db 1769 GTTTGATATTAAATATCGACAGAAATTCATAAAAGAAATTACCTGCTGACCAATTAGG 1828

Qy 565 agtgcactctcttaagaaattacacttctgttacccttggaatccccacacctcgacttt 624

Db 1829 TGTAAAGCGATTGATGATTATACATTAGAGTGAATTAGACAAATCCAGTTCCTTACTT 1888

Qy 625 cttaaaacttttagctcttccagttcttttcccggt---tcataaatctcaagaagacct 681

Db 1889 TGTTCATTTAACAGTTTATCCGGTATTTTACCCTAAATGAAACTTTTCTAAAGCACA 1948

Qy 682 gcaatccaattctctacattatagcagcgagcttctctctcctataataataatacaaaaaa 741

Db 1949 AGGTGATAAGTTTGGGTAGAGCGCAAAATACAAATATATATATGAGGCCATTTGTTATGAG 2008

Qy 742 acaatggaataaactctcaaaaaacccctcactactataatcaaatcaggttggaactaa 801

Db 2009 TGATTGAACACATGAGCAAAAGCTTCCAGTTTAAAGAGAACCCCATCATATTGGGATAATA 2068

Qy 802 aacgattacattacattccattcccgatgcaaacacagcagcaaaactatttaatacagg 861

Db 2069 AACGGTTAAATCGAAGAGATTAACTTT---AATATTGTAAAAAATACGTCACACTGATGT 2125

Qy 862 aaaactcaattggcaggacctctctgggagaagcgcattcctcagaagaacccctatccaa 921

Db 2126 AAATTTATATGACACAAACTCGATTGATGCGGCGGCTCTAACTTCAGAAATTTGTTGATAA 2185

Qy 922 ttacagctotaagggcacttacactcttttgatgctgcaggaacacctatggtcacctt 981

Db 2186 GTTTAGACAAAGCTCAGAAATCCACAGCAAGAAAGAGAGCGGGTTGCTTACCTAAGATT 2245

Qy 982 caatatcaataaattccccctcaacatatagaagcttagagaagccttagcatcagcctt 1041

Db 2246 TAATCAAGTAATCAATCTTAAGTAATAAAACCTTAAGAAAAGCAATTTCAATGTCCTT 2305

Qy 1042 agataaggaagctctcttgctctcaactatatt 1071

Db 2306 TGATCGTGATAACATTGCCGAAAGATTATTTT 2335

Search completed: July 26, 2002, 04:36:20
 Job time: 5828 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:00:52 ; Search time 299.98 Seconds
(without alignments)
10296.450 Million cell updates/sec

Title: US-09-824-567-1
Perfect score: 1799
Sequence: 1 acttccccctgtaacta.....taactgtcttctctctcta 1799

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1799	100.0	1799	22	AAD20238
2	1786.4	99.3	1230025	20	AAX91990
3	449.6	25.0	1038602	20	AAX201425
4	255	14.2	1235	24	AAS18752
5	181.4	10.1	1787	24	AAS18750
6	74.6	4.1	1536	20	AAX61764
7	74.6	4.1	1587	20	AAX61763
8	63.8	3.5	1028	20	AAX20051
9	63.8	3.5	1121	20	AAX13317
					Chlamydia pneumoniae
					Nucleotide sequence
					Complete genome se
					Chlamydia pneumoniae
					Chlamydia pneumoniae
					B. burgdorferi ant
					B. burgdorferi ant
					Enterococcus faeca
					Enterococcus faeca

10	63.8	3.5	1122	20	AAX20050	Enterococcus faeca
11	63.6	3.5	6300	21	AAA51864	B. subtilis opp op
12	62.6	3.5	1564	20	AAX20031	Enterococcus faeca
13	62.6	3.5	1641	20	AAX20030	Enterococcus faeca
14	62.6	3.5	9021	20	AAX13092	Enterococcus faeca
15	60.4	3.4	15747	20	AAX13018	Enterococcus faeca
16	59.6	3.3	1521	20	AAX61568	B. burgdorferi ant
17	59.6	3.3	1572	20	AAX61567	B. burgdorferi ant
18	57.6	3.2	1038602	20	AAX201425	Complete genome se
19	55	3.1	1579	20	AAX20041	Enterococcus faeca
20	55	3.1	1680	20	AAX20040	Enterococcus faeca
21	55	3.1	5504	20	AAX13040	Enterococcus faeca
22	52.2	2.9	26811	20	AAX20253	Borrelia burgdorfe
23	50.8	2.8	716	21	AAA63289	C. trachomatis LGV
24	50.8	2.8	716	22	AAH56193	Chlamydia trachoma
25	49.6	2.8	1597	20	AAX20043	Enterococcus faeca
26	49.6	2.8	1680	20	AAX20042	Enterococcus faeca
27	49.6	2.8	9509	20	AAX13301	Enterococcus faeca
28	48	2.7	1653	23	AAS2765	Enterococcus faeca
29	47.6	2.6	6290	24	ABL33046	Human immune syste
30	47.2	2.6	1545	20	AAX61586	B. burgdorferi ant
31	47.2	2.6	1626	20	AAX61585	B. burgdorferi ant
32	46	2.6	10996	20	AAX13139	Enterococcus faeca
33	46	2.6	18997	24	ABL32570	Human immune syste
34	45.6	2.5	1585	20	AAX20021	Enterococcus faeca
35	45.6	2.5	1659	20	AAX20020	Enterococcus faeca
36	45.6	2.5	11739	20	AAX13087	Enterococcus faeca
37	44.8	2.5	1119	20	AAX20057	Enterococcus faeca
38	44.8	2.5	1242	20	AAX20056	Enterococcus faeca
39	44.8	2.5	2991	20	AAX13168	Enterococcus faeca
40	44.4	2.5	16842	22	AAS46411	Enterococcus faeca
41	44.4	2.5	16842	24	AAS61335	Tumour suppressor
42	44.2	2.5	747	20	AAX99638	Human gene regulat
43	44.2	2.5	7309	22	AAS46568	Nucleic acid seque
44	44.2	2.5	7309	24	ABL33817	Tumour suppressor
45	43.2	2.4	6200	22	AAS46442	Human immune syste
						Tumour suppressor

ALIGNMENTS

RESULT 1

AAD20238

ID AAD20238 standard; DNA; 1799 BP.

XX

AC AAD20238;

XX

XX

DT 15-JAN-2002 (first entry)

XX

DE Chlamydia pneumoniae ATP-binding cassette gene.

XX

XX ATP-binding cassette; antibiotic; vaccine; infection; therapy; poxvirus;

KW ds.

XX

OS Chlamydia pneumoniae.

XX

Key Location/Qualifiers

FT 101..1699

FT /*tag= a

FT /product= "ATP-binding cassette protein"

XX

PN WO200174863-A2.

XX

XX 11-OCT-2001.

PD

PF 04-APR-2001; 2001WO-CA00455.

XX

PR 04-APR-2000; 2000US-194464P.

XX

PA (AVET) AVENTIS PASTEUR LTD.

XX

PI Murdin AD, Oomen RP, Wang J, Dunn P;

XX

DR WPI: 2001-648549/74.
XX P-PSDB; AAE12212.
PT Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for
PT preventing, diagnosing and treating Chlamydia infections in mammals, in
PT particular humans -
XX
XX Claim 2; Fig 1; 8pp; English.
XX
CC The present invention relates to novel Chlamydia pneumoniae ATP-binding
CC cassette protein and its corresponding gene. Sequences of the invention
CC are useful for detecting Chlamydia infection by assaying a body fluid
CC of a mammal with the components. They are also used as vaccines. ATP
CC binding cassette antibodies and vaccines of the invention are useful
CC for preventing or treating Chlamydia infection e.g. infection caused
CC by C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum in mammals,
CC such as humans. The nucleic acid molecules are useful for producing
CC ATP-binding cassettes, in the construction of vaccine vectors such
CC as poxviruses, which are further useful for preventing and/or treating
CC Chlamydia infection and in the construction of attenuated Chlamydia
CC strains that can over-express the nucleic acid molecules or express
CC it in a non-toxic, mutated form. The present sequence is a gene encoding
CC Chlamydia pneumoniae ATP-binding cassette.
XX
SQ Sequence 1799 BP; 560 A; 439 C; 294 G; 506 T; 0 other;

Query Match 100.0%; Score 1799; DB 22; Length 1799;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acttccccctgtataactatgctcagataatgctgtatgattgagctctaggaggag 60
Db 1 acttccccctgtataactatgctcagataatgctgtatgattgagctctaggaggag 60
QY 61 aaaaatttccaaaaaacttagtattccggaattcttatatgcgaagataatcagttgg 120
Db 61 aaaaatttccaaaaaacttagtattccggaattcttatatgcgaagataatcagttgg 120
QY 121 aatctgtatcaccaattctctttagctctccgttagtctccagagctcgaagagttccag 180
Db 121 aatctgtatcaccaattctctttagctctccgttagtctccagagctcgaagagttccag 180
QY 181 tcaactctctacatctcggggagactcgtattataataagaagatgaaccccgcttttt 240
Db 181 tcaactctctacatctcggggagactcgtattataataagaagatgaaccccgcttttt 240
QY 241 agatccagaagaagtgcgactcttttcagaaatcagccttggtcaacatctatgaggg 300
Db 241 agatccagaagaagtgcgactcttttcagaaatcagccttggtcaacatctatgaggg 300
QY 301 attagttcagaagaataattcttccaggaaatagagcctgctcttcgagaagactactc 360
Db 301 attagttcagaagaataattcttccaggaaatagagcctgctcttcgagaagactactc 360
QY 361 tcttctctcggcagactcaactatacttttaactgaatacagctcttttggagtaatgg 420
Db 361 tcttctctcggcagactcaactatacttttaactgaatacagctcttttggagtaatgg 420
QY 421 cgaaccttcaacgctgaagacttttatagaattcttggaacaaagtagctactcaagaagt 480
Db 421 cgaaccttcaacgctgaagacttttatagaattcttggaacaaagtagctactcaagaagt 480
QY 481 ctcagggaatctatgcttttgccttgaattccaattaaaaatgtacgaagaatcccaagggg 540
Db 481 ctcagggaatctatgcttttgccttgaattccaattaaaaatgtacgaagaatcccaagggg 540
QY 541 acaactctccatagaccatttggagtgacactctcotaatgaatcactatgttgttac 600
Db 541 acaactctccatagaccatttggagtgacactctcotaatgaatcactatgttgttac 600
QY 601 cctggaatcccaactcagctatttcttaaaacttttagctcttcagctctttttcccggt 660
Db 601 cctggaatcccaactcagctatttcttaaaacttttagctcttcagctctttttcccggt 660

Db 601 cctggaatcccaactcagctatttcttaaaacttttagctcttcagctctttttcccggt 660
QY 661 tcataaatctcaaaagaacccctgcaatccaaatctctacatctatagcaagcgagctttcta 720
Db 661 tcataaatctcaaaagaacccctgcaatccaaatctctacatctatagcaagcgagctttcta 720
QY 721 tcttaaaaaatcaaaacaaatgataaaactctcaaaaaactcactactataa 780
Db 721 tcttaaaaaatcaaaacaaatgataaaactctcaaaaaactcactactataa 780
QY 781 tcaaatcgagtggaactaaacgattacgattcacttccatcccgatgcaaacacagc 840
Db 781 tcaaatcgagtggaactaaacgattacgattcacttccatcccgatgcaaacacagc 840
QY 841 agcaaaactatttaatacagggaacaaactcaattggcaaggacctctctggggaagcgaat 900
Db 841 agcaaaactatttaatacagggaacaaactcaattggcaaggacctctctggggaagcgaat 900
QY 901 tctcagaagaacctatccaatttcaagtctaaaggcgcaacttcaactcttttgaatgcg 960
Db 901 tctcagaagaacctatccaatttcaagtctaaaggcgcaacttcaactcttttgaatgcg 960
QY 961 aggaacctcagtgctcactcctcaatatacaataaattccccctcaacaatatgaagcttag 1020
Db 961 aggaacctcagtgctcactcctcaatatacaataaattccccctcaacaatatgaagcttag 1020
QY 1021 agaagccttagcctcagccttagataagaagcctcttctcactatattcttaggcg 1080
Db 1021 agaagccttagcctcagccttagataagaagcctcttctcactatattcttaggcg 1080
QY 1081 tgcataaaactcgcgcatctcctcactcaataatctcatagctatcccgaacataaaa 1140
Db 1081 tgcataaaactcgcgcatctcctcactcaataatctcatagctatcccgaacataaaa 1140
QY 1141 acaagagatggaacacgcgcaagcttacgtaaaactctttaaagaagctttagaaga 1200
Db 1141 acaagagatggaacacgcgcaagcttacgtaaaactctttaaagaagctttagaaga 1200
QY 1201 actccaaactcactcgaagatctcgaacatcttaattcttctcccgcttctccgctc 1260
Db 1201 actccaaactcactcgaagatctcgaacatcttaattcttctcccgcttctccgctc 1260
QY 1261 agcaagttcttttagctcctcaacttatcagagacagtggaagaagtttagggctcgc 1320
Db 1261 agcaagttcttttagctcctcaacttatcagagacagtggaagaagtttagggctcgc 1320
QY 1321 tctcctattgtcggaaagaaattgctctctcccaagcagacatctctccagggaaactt 1380
Db 1321 tctcctattgtcggaaagaaattgctctctcccaagcagacatctctccagggaaactt 1380
QY 1381 ctcttagctacagagagtggttcgagactttgctgactcttagtggaattcttaacgat 1440
Db 1381 ctcttagctacagagagtggttcgagactttgctgactcttagtggaattcttaacgat 1440
QY 1441 ctttcttccatccagagtgctcctcttattgcaatcaacataaaggacttctcagaat 1500
Db 1441 ctttcttccatccagagtgctcctcttattgcaatcaacataaaggacttctcagaat 1500
QY 1501 tctcaaaactatagaacagagcgaagatcaacaaacgcctcggaattagtgctcgaagc 1560
Db 1501 tctcaaaactatagaacagagcgaagatcaacaaacgcctcggaattagtgctcgaagc 1560
QY 1561 tctcttttactagagacatttcatattattgagcagatctaccacagcgacttcaatt 1620
Db 1561 tctcttttactagagacatttcatattattgagcagatctaccacagcgacttcaatt 1620
QY 1621 tgcctatgaataaaaaacttttcaatctagggagctcaccacagaggttgagactccg 1680
Db 1621 tgcctatgaataaaaaacttttcaatctagggagctcaccacagaggttgagactccg 1680
QY 1681 ttatgctaaagaaatttagcactcttttaactcgcataacttgcataaactgaactgaatt 1740
Db 1681 ttatgctaaagaaatttagcactcttttaactcgcataacttgcataaactgaactgaatt 1740

QY 1321 tatccctattgtcggaaggaattgtctcttctcccaagcagacctatcttcagggaactt 1380
 Db 245700 tatccctattgtcggaaggaattgtctcttctcccaagcagacctatcttcagggaactt 245759
 QY 1361 ctcttagctacagagagatggttcgagactttgtctgctatgctatgctatttcaagat 1440
 Db 245760 ctcttagctacagagagatggttcgagactttgtctgctatgctatgctatttcaagat 245819
 QY 1441 ctctgtctatccatcagagtgctctctctatgcaatcaacataagagcttcctagaat 1500
 Db 245820 ctctgtctatccatcagagtgctctctctatgcaatcaacataagagcttcctagaat 245879
 QY 1501 tctacaaacatagacaagagcaagatcaccaaaacgctcggaattagtgctgcaagc 1560
 Db 245880 tctacaaacatagacaagagcaagatcaccaaaacgctcggaattagtgctgcaagc 245939
 QY 1561 tctctttacctagagacctctctatattattagccgagctatcacacagcattcaatt 1620
 Db 245940 tctctttacctagagacctctctatattattagccgagctatcacacagcattcaatt 245999
 QY 1621 tctatgaataaaaaactttc-taactagagtgctctcccaaacgctcggaattagtgctgcaagc 1679
 Db 246000 tctatgaataaaaaactttc-taactagagtgctctcccaaacgctcggaattagtgctgcaagc 246059
 QY 1680 gttatcctaaggaataattagcacctcttttaattctgcgaacttgctcaagaactgaattc 1739
 Db 246060 gttatcctaaggaataattagcacctcttttaattctgcgaacttgctcaagaactgaattc 246119
 QY 1740 tatactaaactgggtgctcttggtgagcactcttctctctgactctctctctctcteta 1799
 Db 246120 tatactaaactgggtgctcttggtgagcactcttctctctgactctctctctctcteta 246179

RESULT 3
 AAZ01425/C
 ID AAZ01425 standard; DNA; 1038602 BP.

XX AC AAZ01425;

XX DT 07-OCT-1999 (first entry)

XX DE Complete genome sequence of Chlamydia trachomatis.

XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

XX OS Chlamydia trachomatis.

XX PN WO9928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-IB01939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97FR-0015041.

XX PR 17-DEC-1997; 97FR-0016034.

XX PA (GEST) GENSET.

XX PI Griffais R;

XX DR WPI; 1999-371125/31.

XX PT Genome sequence of Chlamydia trachomatis

XX PS Claim 1; Page 373-656; 1755pp; English.

XX CC The present sequence represents the complete genome of Chlamydia

trachomatis. Open reading frames (ORFs) of the genome encode

CC polypeptides AAV36754-Y37949. The polypeptides can be used as vaccines
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 CC be used to control growth of the microorganism. Chlamydia trachomatis is
 CC responsible for a large number of diseases, e.g. eye diseases such as
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
 CC conjunctivitis; genital diseases such as nongonococcal urethritis;
 CC epididymitis, cervicitis, salpingitis, perihhepatitis, Bartholinitis;
 CC pneumopathy in breast feeding infants; and venereal
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in
 CC treating these diseases.

XX
 SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 25.0%; Score 449.6; DB 20; Length 1038602;
 Best Local Similarity 56.7%; Pred. No. 2.4e-115;
 Matches 976; Conservative 0; Mismatches 696; Indels 50; Gaps 6;

QY 1 acttccccctgctaactatgctcagataatgctgctatgctcaggtctaggggag 60
 Db 252200 ACTTCCCTTCTTCGAGTATGTTTCAAGGAGGTCATTCCATGCGCAAGATATCAGT 252141
 QY 61 aaaaattttcaaaaaactctagttatcc---ggaaattgttatatgcgaagatcagt 117
 Db 252140 GACTGTTTTCGAATCGCACACATGTTTCAAGGAGGTCATTCCATGCGCAAGATATCAGT 252081
 QY 118 gggaaattgttatcacaactctcttagctctcgttagtctctccaaggctcgaaggagtc 177
 Db 252080 GGGAAATGTTGTTCTGCTTCTAGCATAGCAACTTCTGGATGTTCAAAATCTCTCTAACGC 252021
 QY 178 cagtcactctctacatctcctgggagaaactcgtctataataagaagatgaaccccttc 237
 Db 252020 AACCCATCGCTCTCAGCTACTCACACAGTGTCTGTAAGCGTAAAGAAATGATCTCGCAC 251961
 QY 238 tttagatccaagaacaagtgcgactcttttcagaaatcagctctgcaaacatctatga 297
 Db 251960 ATTTGATCTCGAGAGGTTGCGCTTCTTCTGATATCAATTTGATCATCATCTATGA 251901
 QY 298 gggattagttccaagaaaataatcttccaggaaaatatagagctctcttgcgaagacta 357
 Db 251900 AGGATTGGTACAGAA---ACTCCTCTGGAGAAAGTCTTCCCTGCTTTAGCGGAGAGTTT 251844
 QY 358 ctctcttctctcgcgagactcacttatacttttaactgaaatcagctcttttggagtaa 417
 Db 251843 CTCTCTTATCGGAGATATAAAACTTATCTTCACTTCACTTCAAAAGGCTTTTGGAGCAA 251784
 QY 418 tggcagaccttaacagctgaagactttatagaactcttggaaacaagttagctactcaaga 477
 Db 251783 TGGAGATCTTATTACCGCTCATGATTTTGTTCCTGGAATGATGTTTACAAATCG 251724
 QY 478 agtctcaggaatctatgcttctgcttgaaatccaattaaaaatgacgaagatccaaga 537
 Db 251723 TGTGCGTAGTATTTATCTTTCGCCCTTTCCTCTATGA----- 251683
 QY 538 gggacacctctccatagaccatttttggagtgcaactctcttaactgaatcactgtttot 597
 Db 251684 -----CGTGAATAGGATCTGGATTTTTCGCAAGATGATCATCTCTGTAT 251634
 QY 598 taccctggaatcccaacctcgtattctttaaacttttagctctctcagctcttttcc 657
 Db 251633 CAATCTCTCTACTCCAACTCCACATTTTCTNAAGCTGCTTACCCTCCCGTATTTATCC 251574
 QY 658 cgttcat---aaatctcaagaacctgcaatccaactctctactatagcgaagggagc 714
 Db 251573 TGTGCAATTCGAGCATCAGATATGGAAAGAAAGAAATCTCTCCGATATCTACTGGAGC 251514
 QY 715 ttctatctcaaaaatatcaaaaacaaatgagataaaactctcaaaaacccctcacta 774
 Db 251513 TTTTCTTAAAGAGAGAGAAAGACCCGAAGATGGTTAAGCTAGAGAAGAGGCCCTACTA 251454
 QY 775 ctataatcaagtcagtggaactaaaacgattacgattcacttccatcccgatgcaaa 834
 Db 251453 CTATAATAAGAACCCAGGTAGCTGTACAGGAGATCTGTATACACATCATTTCCCTGATACACA 251394

Db 1330 ccaaggcatggtaccactgtcttttaaaagaaacgtcgcaaggagattttctcatagc 1389
 QY 1390 tacagaggaggttgcagaaatttgctgacccctgacatttctaaacgattttgctta 1449
 Db 1390 gacaggagaggttgcagaaatttgctgacccctgacatttctaaacgattttgctta 1449
 QY 1450 tccatcaggaggttgcagaaatttgctgacccctgacatttctaaacgattttgctta 1488
 Db 1450 cccagagacctcacaaatggagaaacagtgattacga 1488

RESULT 6
 AAX61764
 ID AAX61764 standard; DNA; 1536 BP.
 AC AAX61764;
 XX
 DT 19-JUL-1999 (first entry)
 XX
 DE B. burgdorferi antigenic protein coding sequence, t606.nt.
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
 OS Borrelia burgdorferi.
 XX
 PN W0859071-A1.
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US12718.
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (WEDI-) MEDIMMUNE INC.
 XX
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
 DR WPI: 1999-189980/16.
 DR P-PSDB; AAY20067.
 XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases
 PT caused by Borrelia, particularly Lyme disease
 XX
 PS Claim 1; Page 182-183; 275pp; English.
 XX
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.
 XX
 SQ Sequence 1536 BP; 631 A; 246 C; 256 G; 403 T; 0 other;

Query Match 4.1%; Score 74.6; DB 20; Length 1536;
 Best Local Similarity 46.5%; Pred. No. 1.1e-10;
 Matches 602; Conservative 0; Mismatches 639; Indels 54; Gaps 9;

QY 290 atctatgaggagatttcaagaaataatttttcaggaataatagagctgtcttgca 349
 Db 121 atgtttagaggagatttgcagagagacttaacaggggggaaataaacccggacttgca 180
 QY 350 gaagactactcttcttcctcgagagagactcaattatcttttaaaactgaaatcagcttt- 408
 Db 181 aaagggtggagatttcttctgagggagacaggtttacacattttaacctaagagaaatc 240

QY 409 --ttggagtaattggcgacccttaacagctgaagactttatagaattcttggaaacaagta 466
 Db 241 acttggagtgaggttgcagaaatttgctgacccctgacatttctaaacgattttgctta 300
 QY 467 gtaactcaagaagtctcaggaattctgttttgccttgaatcc---aattaaatgta 523
 Db 301 ttaataaagaaactggtcacaagtacgttgaaattggttaaatcggttaataaaatggt 360
 QY 524 cgaagatccaaagagagacacactctccatagaccattttggagtgactctcctaataa 583
 Db 361 caaaaatattttgagggacaagtgaactgactctgaacttggaattggaatgagcgattgaa 420
 QY 584 tctacactgtgttaccctggaatcccaactcgcatttctttaaacttttagctctt 643
 Db 421 aaaaacttagaataaacactggaatcaccacaaacttattttattgtattgtattacac 480
 QY 644 ccagctttttcccggtt-----ataatctcaagaagacccctgcaatcc 688
 Db 481 caatcattttccagttaccagttctgttaccgaaagtatggacaaactggcaagc 540
 QY 689 aaatctctacatagcaagcgaggtttcttcttctaaatatacaaaacaaatgg 748
 Db 541 ccgaaacacatggtgacaagtgtctcttttaataaagaaagaaattcttaacgaaaaa 600
 QY 749 ataaactctcaaaaacccctcactactataatcaaatgctcagtggaataaactgatt 808
 Db 601 tatgtctttgaaaaataaataaataactacactcaatgaatgaatgaagagatt 660
 QY 809 acgattcactctatccgatgcaaacacagcagcaaaactatttaacagggaaactc 868
 Db 661 acattttacacaaatgacagctcaacagcgtataaaattatgaaaaatgaagact- 719
 QY 869 aatggcagagactcctctgggagacgcaattctcagaagaacctatccaattacag 928
 Db 720 -----agatgcaattttgttccatacccccagatctaatcaaaatctaaaa 768
 QY 929 tctaaggggcacttacaactcttttgatgctgcaggaacacctcatgctcactcactatc 988
 Db 769 ttaagaagcgactattactatcactcagctgttaatgccataactttacgcttcaataca 828
 QY 989 aataaattccctcacaacatgaagcttagagaagccttagcatcagccttagaag 1048
 Db 829 cacatcaaacctctgacacgtttaaaatttagaaaagccttaactcttgcattgacaga 888
 QY 1049 gaagctctgtctcaactatattttaggcgtgcgaaacactgcgcatctctactact 1108
 Db 889 gaaacgctta---catataaagtcttgacaacggaactaccctcacaagaagsgcaact 945
 QY 1109 acaaatattcatagctatcccgaaacatcaaaaacagagatggcacaacgcaagcttac 1168
 Db 946 ccaacttttagttcattcttatgcaaaaagttagaattatttattcctgaaatt--- 1002
 QY 1169 gctaaaaaactctttaaagaagcttagaagaactccaatacactgctaaagatctcgaa 1228
 Db 1003 gcaaaacaccttctagctgaagctggata-----tccataaggcaatggattccaa 1054
 QY 1229 catcttaacttattcttccggttctcgtcagcaagtttcttactagtccaactata 1288
 Db 1055 ttttaaaattaaat---acaatacaaacgagcaaaaataaaaaatttgaattatt 1110
 QY 1289 cgagaacagtgtaaaagaagttttaggttcgtctatccctattgtcggaaaggaattgct 1348
 Db 1111 caaaacaaatggaaaaaaatttaaatattgagtgtggaacttgaacaaagaatggaca 1170
 QY 1349 cttctcaagcagacctatctctcaggaacttctttagctacagagaggtggttcga 1408
 Db 1171 acatacttaaacactaaggcgaatgaaattatgaaatgaagacagcaggtatggataggc 1230
 QY 1409 gacttgcgtatcctatgacatttcttaacgatttcttcttactatccat---caggagttcct 1465
 Db 1231 gattatgctatcctttgacatttttaagcatattcaacacaggaatacacacaacttca 1290

QY 431 acagctgaagactttatagaatcttgaaacaaagtagactaactcaagaagtctcgaagc 490
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 341 aaagcaggggatttggtggtggtggtggtggtggtggtggtggtggtggtggtggt 400
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 491 tatgctttgcttgaatccaattaaaattgtacgaagatccaagagggaacacctctcc 550
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 401 agcagtaatacaaatggtatttttaaaatggcggtggtggtggtggtggtggtggtggt 460
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 551 atagaccatttggagtgcaactctcctaataatgaatctacacttgggttgaacctgggaatcc 610
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 461 atggaagaatttgggtgcaagcaatcgatgaccagacactagaaactaacattggaaaaa 520
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 611 ccaacctcgcatctttaaacttttagctctccagttcttttcccggttcataaactc 669
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 521 ccaattcctatttagcccaagctggtggtggagacaccttttatgcctaaaaatgaagc 579
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 RESULT 10
 AAX20050
 ID AAX20050 standard; DNA; 1122 BP.
 XX
 AC AAX20050;
 XX
 DT 20-APR-1999 (first entry)
 XX
 DE Enterococcus faecalis gene EF027.
 XX
 KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic; ss.
 XX
 OS Enterococcus faecalis.
 XX
 PN W09850554-A2.
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08959.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 XX
 DR WPI: 1999-070095/06.
 DR P-PSDB; AAY00060.
 XX
 PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 XX
 PS Claim 1; Page 107; 301pp; English.
 XX
 CC The present sequence represents a gene isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.
 XX
 SQ Sequence 1122 BP; 375 A; 186 C; 263 G; 297 T; 1 other;

Query Match 3.5%; Score 63.8; DB 20; Length 1122;
 Best Local Similarity 50.8%; Pred. No. 1.1e-07;
 Matches 152; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 371 gacggactcacttatacttttaaacatgaatcagcttttttggagtaatgggaacccotta 430
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 280 gacgggttagcttatacattcaagttacgaagcaaaaatggacaacgggatccagtt 339
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 431 acagctgaagactttatagaatcttggaaacaaagtagactactcaagaagtctcaggaatc 490
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 340 aaagcaggggatttggtggtggtggtggtggtggtggtggtggtggtggtggtggt 399
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 491 tatgctttgcttgaatccaattaaaattgtacgaagatccaagagggaacacctctcc 550
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 400 agcagtaatacaaatggtatttttaaaatggcggtggtggtggtggtggtggtggtggt 459
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 551 atagaccatttggagtgcaactctcctaataatgaatctacacttgggttgaacctgggaatcc 610
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 460 atggaagaatttgggtgcaagcaatcgatgaccagacactagaaactaacattggaaaaa 519
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 QY 611 ccaacctcgcatctttaaacttttagctctctccagttcttttcccggttcataaactc 669
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 520 ccaattcctatttagcccaagctggtggtggacaccttttatgcctaaaaatgaagc 578
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
 RESULT 11
 AAA51864
 ID AAA51864 standard; DNA; 6300 BP.
 XX
 AC AAA51864;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE B. subtilis opp operon.
 XX
 KW opp operon; Spook; oligopeptide permease; sporulation; ABC transporter;
 KW ATP-binding cassette transporter; mutation; protein production; oppA;
 KW oppB; oppC; oppD; oppR; ds.
 XX
 OS Bacillus subtilis.
 XX
 FH Location/Qualifiers
 FT CDS 477..2110
 FT /tag= a
 FT /label= oppA
 FT /product= ligand-binding_protein
 FT /transl_except= (pos:2106..2107,aa:Phe)
 FT /note= "this codon contains an apparent 1 base deletion
 FT which alters the reading frame"
 CDS 2222..3157
 FT /tag= b
 FT /label= oppB
 FT /product= membrane_protein
 FT CDS 3161..4078
 FT /tag= c
 FT /label= oppC
 FT /product= membrane_protein
 FT CDS 4083..5159
 FT /tag= d
 FT /label= oppD
 FT /product= ATPase
 FT CDS 5152..6078
 FT /tag= e
 FT /label= oppF
 FT /product= ATPase
 XX
 PN W0200039323-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 21-DEC-1999; 99WO-US31010.
 XX
 PR 24-DEC-1998; 98GB-0028711.
 XX
 PA (GEW) GENENCOR INT INC.
 PA (DIAZ/) DIAZ-TORRES M.

PA* (FERR/) FERRARI E.
 XX Diaz-Torres M, Ferrari E;
 XX WPI: 2000-452412/39.
 DR P-PSDB; AAY97040, AAY97041, AAY97042, AAY97043, AAY97044.
 XX Production of proteins, such as hormones, enzymes, growth factors or
 PT cytokines, in gram-positive microorganisms containing a mutation in at
 PT least one of the genes of the opp operon gene cluster
 XX Disclosure; Fig 1A-M; 32pp; English.
 XX The opp operon of Bacillus (also known as spoOK operon) encodes an
 CC oligopeptide permease that is required for the initiation of sporulation
 CC and the development of genetic competence. The opp operon is a member of
 CC the family of ATP-binding cassette (ABC) transporters involved in the
 CC import or export of oligopeptides from 3-5 amino acids. Bacillus strains
 CC containing a mutation in the opp operon produce more recombinant protein
 CC than the wild-type strain. Therefore, gram-positive microorganisms,
 CC especially Bacillus strains, containing a mutation in at least one of the
 CC genes of the opp operon can be used for heterologous protein production,
 CC especially hormones, enzymes (preferably proteases such as subtilisin),
 CC growth factors or cytokines.
 XX Sequence 6300 BP; 1919 A; 1222 C; 1462 G; 1697 T; 0 other;

Query Match 3.5%; Score 63.6; DB 21; Length 6300;
 Best Local Similarity 43.4%; Pred. No. 2.8e-07;
 Matches 575; Conservative 0; Mismatches 724; Indels 27; Gaps 5;

QY 410 tggagtaagcgaccccttaacagctgaacattatagaatttggaaacaagtagct 469
 DB 804 tggctaatgagacccctgaactgcaacagattttgaattgcttggaaatggcgctt 863
 QY 470 actcaagaagctcaggaatctatcttttgccttgatccaattaaaatgacgaag 529
 DB 864 gaccctaataatgaatcaacatagcttaccagctctactacataaaagtggaagcg 923
 QY 530 atccaagggagacacctccatagaccattttggagctctcctaatgaattctaca 589
 DB 924 gcgaataccggaagggcagctgacgatgagcagtaaaagctgtggaacaaaacg 983
 QY 590 ctgtgtgtaccctgaatcccaacctcgatttttaaaacttttagctcttccagtc 649
 DB 984 ctgaaggttgaaataaacccgactccatttctactgaattaaactgcgtctatacg 1043
 QY 650 ttttccccgttcataa-----atccaaagaaacctgcactccaat----- 692
 DB 1044 tatatgcgatcaataagaaattgcagagaaaaataaaagtggaatacaaaatgcgga 1103
 QY 693 -ctctacctatagcagcgagctttctatctctataaaatacaaaacaaatggata 751
 DB 1104 gatgattatgtatcaaacggcgcttcaaaatgacggcatgggaacacagcggtctatt 1163
 QY 752 aaactctcaaaaacctcactactaatcaaaagtcaggtggaactaaacagattacg 811
 DB 1164 actctcgaaaaaatgacagctattggataaagacaaagtcacaaactgaagaaatcgat 1223
 QY 812 attcaattctcccgatgcaaacacagcagcaaaactatttaatacagggaataactcaat 871
 DB 1224 atggttatgatacaatacaatacagcaactaaataaaatccaaagctgggaacttgat 1283
 QY 872 tggcaaggaacctcttggggaacgacttctctcaagaacacctatccaatttaccgtct 931
 DB 1284 tggccggtgatgcgctcgagcagc---ttccgacagaatccctgcgacctgaaaaa 1340
 QY 932 aaggggcaacttacaactcttttgatgtgcaggaacctcattgctcaccttcaatacaat 991
 DB 1341 gacggttcttactatgtgagccgattgcagggaggttatttggtacaaatccaactgaa 1400
 QY 992 aaattccccctcaacaatatgaagcttagaagacgcttagctcagccttagataagaa 1051

DB 1401 gctaagccattagacacagctcaatattccgttaaagctttaacatatctgtagccgctag 1460
 QY 1052 gctctgtctcaactatattttaggcgctgcaaaaactgcgcatctctctactacata 1111
 DB 1461 tcatgtgttaaaacggttaacgaaggagagcaaatccccgcaatggctgagtgccgct 1520
 QY 1112 aatattcagctatcccgaaacatcaaaaacgaagagatggcacaacgccaagctta---c 1168
 DB 1521 acaatgaaggatttgagatacaaaagaagatacttcaaaagacaatgatgtcaaaaa 1580
 QY 1169 gctaaaaaactctttaagaagctttagaagaactccaatcactgtctaaagatctcgaa 1228
 DB 1581 gcaaaagataacttgaaaaaggcctaaagaataatgggttaagcaaggcatctgattg 1640
 QY 1229 catcttaattctatcttccggttctctgcagcaagttcttactagtcacaactata 1288
 DB 1441 ccaaaatcaaatgtctcaacactgatgcgacacgcaagcaatcgctcaagcagta 1700
 QY 1289 cgagaacagtgaaagaaagtttaggttcgctatccctatttgcgaaaggaaatttgc 1348
 DB 1701 caagaaatggaagaaaaatttaggcgttgattgttgattgataactcagatggaa 1760
 QY 1349 ctctccaaagacaactatcttcaagggaactctctttagctacagaggaggtggtcgca 1408
 DB 1761 gctctattgataagctccacagccaagattatcaaatcgccgtatgggtggtcggc 1820
 QY 1409 gacttgcgtctctatggcaatttctaacgacttttgcctatccatcaggaggtctcct 1468
 DB 1821 gactcaatgatctctatcaacttcttgaattgttccgcgcaaaaacgaggaaataac 1880
 QY 1469 tatgcaatc---aacataaggacttctctagaatttctcaaaaacatagacaagagcaa 1525
 DB 1881 gatacaggtgggaaaaatccagaattcaaaaagcttctgaatcagtcacaaactgaaaca 1940
 QY 1526 gatcccaaaaacgctcggaattagtgctcgcaagttctcttcttacctaggaacttcat 1585
 DB 1941 gataaaaaacaaacgtgcagagctgctgaaagaacgagaaggtattttcattgatgaatg 2000
 QY 1586 attattgagcgcgtctaccacgcgacttttctaatgtctagaataaaaaactttctaat 1645
 DB 2001 cgggttgcccaacttcttctactatcagatactgtggtacaggtgaaacactcaaaagt 2060
 QY 1646 ctaggagctcacaacagaggtgtggaacttcogttatgtctaaaggaaaattagacctc 1705
 DB 2061 gttatcatgcaaggtactgtgaggtttatttttcagaaacgcataattttaataaggctac 2120
 QY 1706 ttttaa 1711
 DB 2121 gctga 2126

RESULT 12
 AAX20031
 ID AAX20031 standard; DNA; 1564 BP.
 XX AAX20031;
 AC AAX20031;
 XX 20-APR-1999 (first entry)
 XX Enterococcus faecalis EF017 gene fragment.
 DE
 DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic; ss.
 XX Enterococcus faecalis.
 OS
 PN W09850554-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08959.
 XX


```

Query Match          3.5%;   Score 62.6;   DB 20;   Length 1641;
Best Local Similarity 44.7%;   Pred. No. 2.9e-07;
Matches 325;   Conservative 0;   Mismatches 399;   Indels 3;   Gaps 2;
QY 398 aaatcaagctttttggagtaatggagaccccttaacagctgaagactttatagaacttgg #57
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 316 aaagatgcaaatggagtaacggcgagcccaatcacagcaaatgatgttttgaataactcttgg 375
 Qy 458 aaacaagtagctactcaagaagctcaggaatctatgcttttgccttggaatcaataaa 517
 Db 376 aagcgacagtgagcccaaaaacagcttcccgcaagctgtattactttgaagggttaaaa 435
 Qy 518 aatgtacgaagatcccaaggggacacccctctccatagaccatttttgagtgcaactctct 577
 Db 436 aattatcgctgtattgtgacgttagcaaatctaaagaagaggttaggggttaacagccatt 495
 Qy 578 aatgaatcacactgtgtttacccctggaatcccaacacccgactgcttcttaaaactttta 637
 Db 496 gatgaccatacttgggaagtagagcaagctactcctatgatttttcaacaattatg 555
 Qy 638 gctcttcacagctttttcccccgtctcaataatctcaagaacccctgcaatcccaatctcta 697
 Db 556 ggggtaccagctttttatcttaaatgaagcatttgcgaaaaacggcgcaaaaactat 615
 Qy 698 cctatagaagcggagctttctctctaaataatacaacaaaacaaatggataaaactc 757
 Db 616 ggtacatcagctgagtcacacactttcaatggccttcacattagaaggttgggatggc 675
 Qy 758 tcaaaaaacccctactactataatcaagtcagggtggaaactaaacagattacagttcac 817
 Db 676 acgaataacttggtcctatgtagaataaaaattttgggtacagcgaattttcg 735
 Qy 818 ttcattcccgatcaaacacagcagcaaaaactatttaacagggaaactcaattggca 876
 Db 736 ctagataaggttgatgcccaagtagttaagaagtcacatactgggaaaaacttttcgaa 795
 Qy 877 aggaacctctgg--ggagaacgcattctcacaagaacccctatccaaattacagttcaag 934
 Db 796 ggggaagaattagatgtgttaaaattcttgagaaattgttgcaagaacgaagcaat 855
 Qy 935 gggcacttacactcttttgcgcaggaacccctcactgctcacccttcaatcaataataaa 994
 Db 856 gcagcttgaaattctgaaattctcgaacgattatatacccaataataacgcaaaaa 915
 Qy 995 ttcccctcaacatatgagcttagaagccttagcagccttagcagccttagaaggaagct 1054
 Db 916 gatcttttggcaataagaatgcagctcgagcaatagcattatcattgaattctgagcgt 975
 Qy 1055 ctgtctcaactatattctttagccgtgcacaaaactgcgcgactcactcctaccatacaat 1114
 Db 976 ttagtcaaaatgttttaaatgatggtctcaaaaaagcaacttgcttgccacaggt 1035
 Qy 1115 attcata 1121
 Db 1036 ttactata 1042

RESULT 14

AA13092
 ID AAX13092 standard; DNA; 9021 BP.

XX AAX13092;

XX AAX13092;

XX 19-MAR-1999 (first entry)

XX Enterococcus faecalis genome contig SEQ ID NO:155.

XX Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

XX WO98050555-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-US08985.

XX 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Barash SC, Dillon PJ, Kunsch CA;
 PI WPI; 1999-045171/04.
 DR New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 XX Claim 1; Page 880-885; 2084pp; English.
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 XX Sequence 9021 BP; 2967 A; 1582 C; 1878 G; 2581 T; 13 other;
 SQ

Query Match 3.5%; Score 62.6; DB 20; Length 9021;

Best Local Similarity 44.7%; Pred. No. 6.3e-07;

Matches 325; Conservative 0; Mismatches 399; Indels 3; Gaps 2;

Qy 398 aaatcagcttttttgagtaaatggcgaccccttaacagctgaagcattttagaatacttgg 457

Db 3376 aaagatgcaaatggatgaacggcgagcccaatcacagcaaatgatttgaatactctgg 3435

Qy 458 aaacaagtagtaactcaagaagctcaggaatctatgcttttgccttgaatccaataaa 517

Db 3436 aagcgacagtgagcccaaaaacagcttcccgcaagctattactttgaagggttaaaa 3495

Qy 518 aatgtacgaagaatcccaagaggagacacctctccatagacacatttggagtgcaactctct 577

Db 3496 aattatcgctgtattgttgacgttagcaaatctaaagaagaggttaggggttaacagccatt 3555

Qy 578 aatgaatctacactgtttgttacccctggaaaccccaacctgcatttcttaaacatttta 637

Db 3556 gatgaccatacttgggaagtagagcgaatcctctatgagttattttcaacattattg 3615

Qy 638 gctcttcagcttttttcccccgttcaataatctcaagaacccctgcaatcccaatctcta 697

Db 3616 gcggtaccagctttttatctcttaaatgaagcatttgcgaaaaaacggcgcaaaactat 3675

Qy 698 cctatagaagcggagctttctctctcaaaaatacaacaaaacaaatggatgataaactc 757

Db 3676 ggtacatacagctgagtcacacactttcaatggccttcacattagaaggttggatggc 3735

Qy 758 tcaaaaaacccctactactataatcaagtcagggtggaaactaaacgaattacagcgaattctcg 3795

Db 3736 acgaataacttggctcctatgtagaataaaaaattttgggatcagcgaattttcg 3955

Qy 818 ttattcccgatgcaaacacagcagcaaaaactatttaacagggaaaaactcaattggca 876

Db 3796 ctagataaggttggatgcccaagtagttaagaagtcacatactgggaaaaacttttcgaa 3855

Qy 877 aggaacctctctgg--ggagaacgcattctcacaagaacccctatcccaattacagttcaag 934

Db 3856 ggggaagaattagatgttgtaaaaatttcttgagaaatttcttgcaagaacgaaggaat 3915

QY 935 gggcaacttacactcttttgatgocgaggaacotcattgctcacotccaataatcaataa 994
 Db 3916 gcagctttgaaattcgtgaaattcctggaacgtattatataccaattaaacacaa 3975
 QY 995 tccccctcaacaataatgagcgttagagaaagccttagcatcagccttagataaagagct 1054
 Db 3976 gatcttttgcaataaagaatgcagctcgagcaatagcattatcatatgagcgt 4035
 QY 1055 ctgtctcaactatattttaggcgctgcgcaaaactgcgacatcctcactacataa 1114
 Db 4036 ttagctaaatgctttaaataatgatggctcaaaaaagcacttggcttctgccaacaggt 4095
 QY 1115 attcata 1121
 Db 4096 ttcacta 4102

RESULT 15

AAx13018
 ID AAX13018 standard; DNA; 15747 BP.

XX AAX13018;

AC

XX

DT 19-MAR-1999 (first entry)

XX

DE Enterococcus faecalis genome contig SEQ ID NO:81.

XX

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

XX

KW vaccine; attenuation; computer readable medium; ds.

XX

OS Enterococcus faecalis.

XX

PN WO9805055-A2.

XX

PD 12-NOV-1998.

XX

PF 04-MAY-1998; 98WO-US08985.

XX

PR 14-NOV-1997; 97US-0066009.

XX

PR 06-MAY-1997; 97US-0044031.

XX

PR 16-MAY-1997; 97US-0046655.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Dillon PJ, Kunsch CA;

XX

XX WPI; 1999-045171/04.

XX

XX New isolated Enterococcus faecalis polynucleotides and polypeptides

XX

XX - used to develop products for the detection of Enterococcus and for

XX use in vaccines for prevention or attenuation of Enterococcus

XX infection.

XX

XX Claim 1; Page 562-570; 2084pp; English.

XX

XX A computer readable medium has been developed which has recorded on it

XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

XX AAX12938 to AAX13919 represent these nucleotide sequences which are

XX primary nucleotide sequences, also known as contigs. The computer-based

XX system can identify fragments of the Enterococcus faecalis genome with

XX commercial importance. The products can be used to detect the presence

XX of Enterococcus faecalis in samples. They can also be used for

XX diagnosing Enterococcal infection in an animal and monitoring

XX progression of disease, and for identifying agents which can be used to

XX modulate the growth or pathogenicity of Enterococcus faecalis, or

XX another related organism, in vivo or in vitro. In particular the

XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences

XX can be used in vaccines to prevent or attenuate an Enterococcal

XX infection.

XX

XX Sequence 15747 BP; 4949 A; 3274 C; 2618 G; 4899 T; 7 other;

XX

Query Match

Best Local Similarity 3.4%; Score 60.4; DB 20; Length 15747;
 Matches 194; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

QY 273 tcagccttgtoaacatctatctatgagggatagttcgaagaaataatcttctcaggaata 332
 Db 9377 ttaccactttaataatgtctacgaagaaatctctgttagtaagaagcaacaacocg 9436
 QY 333 tagagcctgctcttgcaagactactctcttctctcgagcactcacttatactttta 392
 Db 9437 cgcctgctggtgcagcgaagaaagcactgttttcagaagcgggttttagttacaagtta 9496
 QY 393 a---actgaaatcagcttttttgagtaatgagcccttaacagctgagactttatag 449
 Db 9497 aattacgtgaagaatcaaaatggtctgatggcaaacacagttactgtcgattacgttt 9556
 QY 450 aatcttggaacaagtagctactcaagaagctctcaggaatctatgtcttttgcttgaatc 509
 Db 9557 acggttggaacaagtagctactcaagaagctctcaggaatctatgtcttttgcttgaatc 9616
 QY 510 caattaaaaatgtacgaaagatcccaagggagggacacacctctccatagaccattttgagtc 569
 Db 9617 cagtataaaatgtcgaagaaatttctaaaggggaactacctaagaagaagtgggcatta 9676
 QY 570 actctcctaatgaatctacactgtttgttaccctgggaatcccccaactcgtgattcttaa 629
 Db 9677 aagcaatcaatgatcatgaattagaatcactctagaacacagcaacacacattttgaog 9736
 QY 630 aacttttagctcttcacgtcttttcccccgttctataa 667
 Db 9737 attattggctttccctcttcttcttgcgcgaagctcaa 9774

Search completed: July 26, 2002, 05:01:42

Job time: 7250 sec